

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using bw model

Run on: September 10, 2005, 05:44:18 ; Search time 475 Seconds  
(without alignments)  
9346.960 Million cell updates/sec

Title: US-10-620-487-1

Perfect score: 750  
Sequence: 1 gatccataaccaccaacta.....ttagtaacttgctactaag 750

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_16Dec04:\*

1: geneseqn1980s:\*

2: geneseqn1990s:\*

3: geneseqn2000s:\*

4: geneseqn2001as:\*

5: geneseqn2001bs:\*

6: geneseqn2002as:\*

7: geneseqn2002bs:\*

8: geneseqn2003as:\*

9: geneseqn2003bs:\*

10: geneseqn2003cs:\*

11: geneseqn2003ds:\*

12: geneseqn2004as:\*

13: geneseqn2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	750	100.0	750	12	ADK66005	Adk66005 E.coli yf
2	588	78.4	588	3	AAAs52690	Aaas52690 Escherich
3	295.8	39.4	1668	5	AAS81979	Aas81979 DNA encod
4	265.8	35.4	618	11	ACH99695	Ach99695 Klebsiell
5	156	20.8	372	11	ACH99702	Ach99702 Klebsiell
6	100	13.3	100	8	ACD79726	Acdd79726 E. coli K
7	100	13.3	100	8	ACD79725	Acdd79725 E. coli K
8	100	13.3	100	8	ACD79727	Acdd79727 E. coli K
9	50.4	6.7	573	6	ABK73143	Abk73143 Bacillus
10	42.6	5.7	2325	5	AAS92926	Aas92926 DNA encod
11	42.6	5.7	3020	5	AAS88917	Aas88917 DNA encod
12	42	5.6	669	9	ADA31058	Ada31058 DNA encod
13	39.8	5.3	636	10	ADF02251	Adf02251 Bacterial
14	39.6	5.3	750	11	ABD10542	Abd10542 Pseudomon
15	39	5.2	247682	12	ADL08109	Adl08109 Human gen
16	38.2	5.1	636	11	ACH97476	Ach97476 Klebsiell
17	37.2	5.0	609	10	ACF70009	Acf70009 Photorhab
18	37.2	5.0	110000	10	ACF65385_1	Continuation (2 of
19	37.2	5.0	110000	10	ACF67367_30	Continuation (31 o
20	36.4	4.9	2000	8	ADA71938	Ada71938 Rice gene

C	21	36	4.8	486	13	ADR62125	Adr62125 Cotton cD
C	22	35.6	4.7	492	10	ACD96860	Acdd96860 Human col
C	23	35.6	4.7	612	10	ACF71507	Acf71507 Photorhab
C	24	35.6	4.7	110000	10	ACF67367_46	Continuation (47 o
C	25	35.6	4.7	110000	10	ACF65388_01	Continuation (2 of
C	26	35.2	4.7	474	12	ADI45188	Adi45188 Rice isop
C	27	35.2	4.7	1302	4	ABL21095	Abi21095 Drosophil
C	28	35.2	4.7	79467	9	ADA02717	Ada02717 Mouse Nfa
C	29	35.2	4.7	79467	10	ADB72455	Abd72455 Mouse Nfa
C	30	35.2	4.7	79467	10	ADE95965	Ade95965 Mouse Nfa
C	31	35	4.7	1877	4	ABL14601	Abi14601 Drosophil
C	32	35	4.7	4532	4	ABL14600	Abi14600 Drosophil
C	33	35	4.7	173564	13	ABD32953	Abd32953 Human can
C	34	34.6	4.6	214	3	ADF57362	Adf57362 Urogenita
C	35	34.6	4.6	3821	5	ADL62643	Adl62643 Human ova
C	36	34.6	4.6	3825	5	ABV25634	Abv25634 Human pro
C	37	34.4	4.6	6996	4	ABL21622	Abi21622 Drosophil
C	38	34.2	4.6	592	6	ABT11039	Abt11039 Human bre
C	39	34.2	4.6	645	11	ACH96827	Ach96827 Klebsiell
C	40	34.2	4.6	1707	6	AAS15596	Aas15596 DNA encod
C	41	34.2	4.6	2000	8	ADA71938	Ada71938 Rice gene
C	42	34.2	4.6	2000	12	ADJ41214	Adj41214 Plant cDN
C	43	34.2	4.6	141463	11	ACN43862	Acn43862 Human gen
C	44	33.4	4.5	202	3	AAC25780	Aac25780 Human sec
C	45	33.4	4.5	1055	3	AAA07254	Aaa07254 Asparagin

ALIGNMENTS

RESULT 1	
ADK66005	
ID	ADK66005 standard; DNA; 750 BP.
XX	
AC	ADK66005;
XX	
DT	06-MAY-2004 (first entry)
XX	
DE	E coli yfik gene.
XX	
KW	ds; gene; yfik; phosphoglycerate amino acid; amino acid production.
XX	
OS	Escherichia coli.
XX	
FH	Key
FT	CDS
FT	Location/Qualifiers
FT	110..697
FT	/*tag= a
XX	/product= "yfik"
PN	
XX	EP1382684-A1.
XX	
PD	21-JAN-2004.
XX	
PF	10-JUL-2003; 2003EP-00015546.
XX	
PR	19-JUL-2002; 2002DE-01032930.
XX	
PA	(CONE ) CONSORTIUM ELEKTROCHEM IND GMBH.
XX	
PI	Maier T;
XX	
DR	WPI; 2004-192974/19.
DR	P-PSDB; ADK66006.
XX	
PT	New microbial strain, useful for producing amino acids of the
PT	phosphoglycerate family, particularly serine and cysteine, has increased
PT	activity of the yfik gene product.
XX	
PS	Disclosure; Page 8-10; 16pp; German.
XX	
CC	The present invention relates to a microbial strain for fermentative
CC	product of amino acids of the phosphoglycerate family, or their
CC	derivatives, has higher activity of the product of the yfik gene, or its

CC homologue, compared with its parent strain. The microbial strain is used  
CC to produce O- and N-acetyl-L-serine, L-cysteine and cystine. The present  
CC sequence is the *E. coli* yfik gene.

SQ Sequence 750 BP; 172 A; 169 C; 194 G; 215 T; 0 U; 0 Other;

Query Match	100.0%;	Score 750;	DB 12;	Length 750;
Best Local Similarity	100.0%;	Pred. No. 1.5e-247;		
Matches 750; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

QY	1	GATCCATAACCCCAACCCATATCGAAATAATCGAATCTAGAATATATAAAACATTTCAATTTT	60
Db	1	GATCCATAACCCCAACCCATATCGAAATAATCGAATCTAGAATATATAAAACATTTCAATTTT	60
QY	61	TTAAATGTTCCGTCGCGGTACTGTCTACCAAAACAGAGAGATAACAGTGACACCGAC	120
Db	61	TTAAATGTTCCGTCGCGGTACTGTCTACCAAAACAGAGAGATAACAGTGACACCGAC	120
QY	121	CCTTTAAGTGCCTTTTGGACTTACACCCGTATTACCGCTATGACGCCAGACCGAACA	180
Db	121	CCTTTAAGTGCCTTTTGGACTTACACCCGTATTACCGCTATGACGCCAGACCGAACA	180
QY	181	TATTCTCGCCCTTAGCTCTGCTACGTCGCATGAGATTTGTCAAAGTACCCGCGTGGC	240
Db	181	TATTCTCGCCCTTAGCTCTGCTACGTCGCATGAGATTTGTCAAAGTACCCGCGTGGC	240
QY	241	AGGGATGAGTCTGGGATTTTGATTTGTATGTACTGTGTGCGGGCATTTCAATTTCACT	300
Db	241	AGGGATGAGTCTGGGATTTTGATTTGTATGTACTGTGTGCGGGCATTTCAATTTCACT	300
QY	301	GGCAGTGATGACCCCGGCAGCGGTACACCTTTGAGTTGGCGGGCGGCATATATTGT	360
Db	301	GGCAGTGATGACCCCGGCAGCGGTACACCTTTGAGTTGGCGGGCGGCATATATTGT	360
QY	361	CTGCGTGGCGTGAAAAATCGCCACAGCCCAACAAGGAAGACGACTTCAGGCAAAACC	420
Db	361	CTGCGTGGCGTGAAAAATCGCCACAGCCCAACAAGGAAGACGACTTCAGGCAAAACC	420
QY	421	AATCAGCTTTTGGGCCAGCTTTCCTTGACGTTGTGAACGTCAAAATCATTTTGTACGG	480
Db	421	AATCAGCTTTTGGGCCAGCTTTCCTTGACGTTGTGAACGTCAAAATCATTTTGTACGG	480
QY	481	TGTTACGGCACTGTGCACGTTGTCTGCGCCAAACACAGGCGTTAAGCTGGGTAGTTGG	540
Db	481	TGTTACGGCACTGTGCACGTTGTCTGCGCCAAACACAGGCGTTAAGCTGGGTAGTTGG	540
QY	541	CGTCAGCGTTTGTCTGGCGATGATTGGGACGTTTGGCAATGTGTCTGGGCGTGGCGGG	600
Db	541	CGTCAGCGTTTGTCTGGCGATGATTGGGACGTTTGGCAATGTGTCTGGGCGTGGCGGG	600
QY	601	GCACTGTTCAGCGATTTGTTGCCAGTATGTCGCCAGTTAAATATCGTCTTGCCCT	660
Db	601	GCACTGTTCAGCGATTTGTTGCCAGTATGTCGCCAGTTAAATATCGTCTTGCCCT	660
QY	661	GTTGCTGCTATATTGCCGCGTACGCAATTTTCTATTACGAAAAAAGCGGAAGAGTCCG	720
Db	661	GTTGCTGCTATATTGCCGCGTACGCAATTTTCTATTACGAAAAAAGCGGAAGAGTCCG	720
QY	721	CCTCTTCGCTTAGTAACCTTGCTACTTAAG 750	
Db	721	CCTCTTCGCTTAGTAACCTTGCTACTTAAG 750	

KW E. coli; yfiK gene; amino acid production; excretion protein gene;  
 KW amino acid excretion protein; ds.  
 XX  
 OS Escherichia coli.

Key	Location/Qualifiers
FT	1. .588
FT	/*tag= a
FT	/product= "yfik"
FT	

PN EP1016710-A2.

PD 05-JUL-2000.

PF 17-DEC-1999; 99EP-00125263.

PR 30-DEC-1998; 98RU-00124016.

XX



PI Tokhmakova IL;

DR WPI; 2000-414802/36.



PT Increased production of L-amino acids by an *Escherichia bacterium*  
PT comprises increasing the expression amount of an L-amino acid excretion  
PT protein.

PS Disclosure; Page 22; 29pp; English.

CC The present sequence is the yf1k gene (an excretion protein gene) of  
CC *Escherichia coli*. The amino acid excretion protein produced from this  
CC gene is involved in the production of amino acids, and an increase in its  
CC expression leads to an increased accumulation of amino acids in the cell.  
CC In this case, an increase in threonine, histidine, glutamic acid and  
CC proline is achieved if multiple copies of the gene are transfected into a  
CC bacterium. The bacterium used is *E. coli*

SQ Sequence 588 BP; 113 A; 134 C; 167 G; 174 T; 0 U; 0 Other;

Query Match	78.4%;	Score 588;	DB 3;	Length 588;
Best Local Similarity	100.0%;	Pred. No. 1.1e-191;		
Matches 588;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY 110 GTGACACCGACCCCTTTTAAGTCTTTTGACTTACACCGTGATTACCGCTATGACGCCA 169  
|||||  
Db 1 GTGACACCGACCCCTTTTAAGTCTTTTGACTTACACCGTGATTACCGCTATGACGCCA 60

QY 170 GGACCGACAATATTCTCGCCCTTAGCTCTGCTACGTGCAATGATTTCGTCAAAGTACC 229

Dp 61 GGACCGACAATATTCTCGCCCTTAGCTCTGCTACGTGCAATGATTTCGTCAAAGTACC 120

QY 230 CGCGTCTGCAGGGATGAGTCTGGGATTTTGTATTGTATCTGTGTGCGGGCATT 289  
 |||||  
 Db 121 CGCGTCTGCAGGGATGAGTCTGGGATTTTGTATTGTATCTGTGTGCGGGCATT 180

QY 290 TCATTTCACTGGCAGTGATTGACCCGGCAGCGGTACACCTTTTGA GTTGGCGGGGGCG 349

Dh 181 TCAATTTTCAC TGAGCAGTGA TTGCACCCGGCAGCGGTACA CTTTTTGA GTTGGCGGGGGCG 240

350 GCATATATTGCTGGCTGGCGTGAAAAATGCCACCAGCCCAACAAGGAGACGGACTTT 409

Qy 410 CAGGCAAAACCAATCAGCTTTTGGGCCAGCTTTGCTTTGCAATTGTGAACGTCAAATC 465

470 ATTTTGACGGTGTACGGCACTGTCAGCTTGTCTGCGCAACACAGGCGTTAGC 529

Db 361 ATTTGTACGGTGTTCACGGCACTGTCAGCGTTGTTGTCGCCCAACACAGGCCGTTAAGC 420  
QY 530 TGGGTAGTTGGCGTCAGCGTTTTTGTCTGGCGATGATGGAGCGTTGGCAATGTGCTGG 589  
Db 421 TGGGTAGTTGGCGTCAGCGTTTTGCTGGCGATGATGGAGCGTTGGCAATGTGCTGG 480  
QY 590 GCGCTGGCGGGGCATCTGTTTCAGCGATTGTTTCGCCAGTATGTCGCCAGTTAAATATC 649  
Db 481 GCGCTGGCGGGGCATCTGTTTCAGCGATTGTTTCGCCAGTATGTCGCCAGTTAAATATC 540  
QY 650 GTGCTTGCCCTGTGCTGCTATTTGCGCGGTACGCAATTTCTATTAA 697  
Db 541 GTGCTTGCCCTGTGCTGCTATTTGCGCGGTACGCAATTTCTATTAA 588  
RESULT 3  
AAS81979  
ID AAS81979 standard; cDNA; 1668 BP.  
XX  
AC AAS81979;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #17783.  
XX  
KM Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KM food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US008631.  
XX  
PR 31-MAR-2000; 2000US-00540217.  
PR 23-AUG-2000; 2000US-00649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
DR P-PSDB; ABG17792.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.  
XX  
PS Claim 1; SEQ ID NO 17783; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic  
CC coding sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at

CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 1668 BP; 376 A; 420 C; 445 G; 427 T; 0 U; 0 Other;  
Query Match 39.4%; Score 295.8; DB 5; Length 1668;  
Best Local Similarity 99.3%; Pred. No. 1.5e-90;  
Matches 297; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 382 CACCAGCCCAAGAAAGAACGAGCTTCAGGCAAAACCAATCAGCTTTGGCCAGCTT 441  
Db 1014 CACTCGCCCAACAAAGAAAGACGAGCTTCAGGCAAAACCAATCAGCTTTGGCCAGCTT 1073  
QY 442 TGCTTTGCACTTTGTGAACGTCAAAATCATTTTGTACGGTGTACGGCACTGTGACGTT 501  
Db 1074 TGCTTTGCACTTTGTGAACGTCAAAATCATTTTGTACGGTGTACGGCACTGTGACGTT 1133  
QY 502 TGTTCTGCCGCAACACAGGCGTTAAGCTGGGTAGTTGGCGTCAGCGTTTGTCTGGCGAT 561  
Db 1134 TGTTCTGCCGCAACACAGGCGTTAAGCTGGGTAGTTGGCGTCAGCGTTTGTCTGGCGAT 1193  
QY 562 GATTGGACGTTTGGCAATGTGTGCTGGCGCTGGCGGGGCATCTGTTTCAGCGATTGTT 621  
Db 1194 GATTGGACGTTTGGCAATGTGTGCTGGCGCTGGCGGGGCATCTGTTTCAGCGATTGTT 1253  
QY 622 TCGCCAGTATGTCGCCAGTTAAATATCGTCTTGCCCTGTGCTGCTAATTGCGCGG 680  
Db 1254 TCGCCAGTATGTCGCCAGTTAAATATCGTCTTGCCCTGTGCTGCTAATTGCGCGG 1312  
RESULT 4  
ACH99695  
ID ACH99695 standard; DNA; 618 BP.  
XX  
AC ACH99695;  
XX  
DT 29-JUL-2004 (first entry)  
XX  
DE Klebsiella pneumoniae polynucleotide seqid 5490.  
XX  
KM Recombinant expression vector; transcription regulatory element;  
KM Klebsiella pneumoniae protein; antibacterial; Vaccine; gene; ds.  
XX  
OS Klebsiella pneumoniae.  
XX  
PN US6610836-B1.  
XX  
PD 26-AUG-2003.  
XX  
PF 27-JAN-2000; 2000US-00489039.  
XX  
PR 29-JAN-1999; 99US-0117747P.  
XX  
PA (GENO-) GENOME THERAPEUTICS CORP.  
XX  
PI Breton GL, Osborne M;  
XX  
DR WPI; 2003-895346/82.  
DR P-PSDB; ABO66144.  
XX  
PT New nucleic acid encoding a Klebsiella pneumoniae polypeptide, useful for  
PT preparing a vaccine composition against Klebsiella pneumoniae.  
XX  
PS Disclosure; SEQ ID NO 5490; 932pp; English.  
XX  
CC The invention describes a new isolated nucleic acid encoding a Klebsiella  
CC pneumoniae polypeptide. Also described are: a recombinant expression  
CC vector comprising the nucleic acid, operably linked to a transcription  
CC regulatory element; and a cell comprising the recombinant expression  
CC vector. The nucleic acid is useful for preparing a vaccine composition  
CC against Klebsiella pneumoniae. This sequence encodes a Klebsiella  
CC pneumoniae polypeptide of the invention  
XX  
SQ Sequence 618 BP; 110 A; 161 C; 183 G; 164 T; 0 U; 0 Other;



Query Match	35.4%	Score 265.8;	DB 11;	Length 618;
Best Local Similarity	65.0%;	Pred. No. 1.9e-80;		
Matches 393; Conservative	0;	Mismatches 212;	Indels 0;	Gaps 0;

[illegible]

Accession	Source	Organism	Accession	Source	Organism
ACH99702	standard, DNA, 372 BP.		ACH99702	standard, DNA, 372 BP.	
ACH99702;			ACH99702;		
29-JUL-2004	(first entry)		29-JUL-2004	(first entry)	
Klebsiella pneumoniae polynucleotide seqid 5497.			Klebsiella pneumoniae polynucleotide seqid 5497.		
Recombinant expression vector; transcription regulatory element;			Recombinant expression vector; transcription regulatory element;		
Klebsiella pneumoniae protein; antibacterial; Vaccine; gene; ds.			Klebsiella pneumoniae protein; antibacterial; Vaccine; gene; ds.		
Klebsiella pneumoniae.			Klebsiella pneumoniae.		
US6610836-B1.			US6610836-B1.		
26-AUG-2003.			26-AUG-2003.		
27-JAN-2000; 2000US-00489039.			27-JAN-2000; 2000US-00489039.		
99US-0117747P.			99US-0117747P.		

PA (GENO-) GENOME THERAPEUTICS CORP.  
XX  
XX Breton GL, Osborne M;  
XX  
XX WPI; 2003-895346/82.  
DR P-PSDB; ABO66151.  
XX  
XX  
PT New nucleic acid encoding a *Klebsiella pneumoniae* polypeptide, useful for  
PT preparing a vaccine composition against *Klebsiella pneumoniae*.  
XX  
XX  
PS Disclosure; SEQ ID NO 5497; 932pp; English.  
XX  
XX The invention describes a new isolated nucleic acid encoding a *Klebsiella*  
CC *pneumoniae* polypeptide. Also described are: a recombinant expression  
CC vector comprising the nucleic acid, operably linked to a transcription  
CC regulatory element; and a cell comprising the recombinant expression  
CC vector. The nucleic acid is useful for preparing a vaccine composition  
CC against *Klebsiella pneumoniae*. This sequence encodes a *Klebsiella*  
XX *pneumoniae* polypeptide of the invention  
XX  
SQ Sequence 372 BP; 87 A; 95 C; 87 G; 103 T; 0 U; 0 Other;

Query Match	20.8%	Score 156;	DB 11;	Length 372;
Best Local Similarity	71.8%;	Pred. No. 1.1e-42;		
Matches 204; Conservative	0;	Mismatches 80;	Indels 0;	Gaps 0

QY	19	TATCGAAAATATCGAATCTAGAATATAAAAA	CATTCA	TTTTTTTAAATG	TTCCGTGCG	78
Db	85	TATCGATATTATCGAAGCAGAA	TGA	AAAAATATTCAT	TTTTTTTAAATATACGCTTCGG	144
QY	79	GTACTGTCTACCAAAACAGAGGAGATAACA	AGTGACA	CCGACCTTTTAA	GTGCTTTTG	138
Db	145	GTAGTGTCTGCCAGACACA	GAGAGAAAACCGGTGACACCTAC	CCGTGATAGCGCTTTT		204
QY	139	GACTTACACCCCTGATTACCGCTATGAC	CCGACGACCCGAACA	TATTTCTG	CCCTTAGCTC	198
Db	205	AACTTACACTTTGATCATCCGCCCTGAC	ACCCGGGCCCAACA	TATCCTGCGCTGACGACAG		264
QY	199	TGCTACGTCGCATGGA	TTTCGTCAAAGTACCCCGCTGCTGGCAGGGA	TGAGTTCGGAA	TT	258
Db	265	CGTCACCTCCCATGGTCTACGCCGACGCTG	CGGGTGCTGGCCGGGATGAGCGT	CGGTT		324
QY	259	TTTGATTGTGATGTTACTGTGTGCGGGCAT	TTTCATTTTCACTGG		302	
Db	325	TATTATCACTATGTGATTGGCGCGCTT	TAACCTTTTTCGCTGG		368	

RESULT 6	
ACD79726	
ID	ACD79726 standard; DNA; 100 BP.
XX	
AC	ACD79726;
XX	
DT	19-SEP-2003 (first entry)
XX	
DE	E. coli K12 MG1655 biochip probe SEQ ID 11002.
XX	
KM	Biochip; gene expression; gut; diagnostic; detection; probe; ss.
XX	
OS	Escherichia coli.
XX	
PN	EP1260592-A1.
XX	
PD	27-NOV-2002.
XX	
PF	17-MAY-2001; 2001EP-00112179.
XX	
PR	17-MAY-2001; 2001EP-00112179.
XX	
PA	(MMGB-) MMG-BIOTECH AG.
XX	
PI	Donner H, Drescher B, Huber A, Weber J;
XX	



DR WPI; 2003-241155/24.  
XX  
PT Biochip containing probes complementary with open reading frames in  
PT Escherichia coli K12, useful for detecting gene expression and expression  
PT patterns.  
XX  
PS Claim 3; Page 1714; 2004pp; German.  
XX  
CC This invention describes a novel biochip comprising probe spots, each  
CC containing many identical probes. The probes are nucleotide sequences of  
CC 30-80 bases, are prepared ex situ from synthetic oligonucleotides and at  
CC least one includes a segment of at least 20 bases identical with, or  
CC complementary to, a segment of an open reading frame (orf) of Escherichia  
CC coli K12. The biochip is used for specific detection of gene expression  
CC in K12 and for determining the gene expression pattern, e.g. for  
CC diagnostic determination of which E. coli strains are present in the gut,  
CC and to determine the effects of e.g. growth media on gene expression. The  
CC biochip provides as comprehensive as possible detection of the K12  
CC genome, with simultaneous analysis of many different genes with a single  
CC device, and comparison of gene expression between K12 and its mutants or  
CC other E. coli strains in a single experiment. Apart from qualitative and  
CC quantitative information about gene expression, it also allows  
CC measurements of population densities for the various strains. The use of  
CC synthetic oligonucleotides for preparation of probes allows free  
CC variation in probe length and ensures high purity (and thus selectivity,  
CC reactivity and reproducibility); also synthetic probes are generally  
CC shorter than probes prepared by polymerase chain reaction. ACD68731 to  
CC ACD81540 represent oligonucleotide probes used with the biochip described  
CC in the invention  
XX  
SQ Sequence 100 BP; 18 A; 20 C; 30 G; 32 T; 0 U; 0 Other;  
  
Query Match 13.3%; Score 100; DB 8; Length 100;  
Best Local Similarity 100.0%; Pred. No. 1.1e-23;  
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 466 AATCATTTGTACGGTGTACGGCACTGTGCAGTTGTTCTGCCGCAACACAGGCGTT 525  
|||||  
Db 1 AATCATTTGTACGGTGTACGGCACTGTGCAGTTGTTCTGCCGCAACACAGGCGTT 60  
  
QY 526 AAGCTGGTAGTGGCGTCAGCGTTTGTCTGGCGATGATT 565  
|||||  
Db 61 AAGCTGGTAGTGGCGTCAGCGTTTGTCTGGCGATGATT 100  
  
RESULT 7  
ACD79725  
ID ACD79725 standard; DNA; 100 BP.  
XX  
AC ACD79725;  
XX  
DT 19-SEP-2003 (first entry)  
XX  
DE E. coli K12 MGI655 biochip probe SEQ ID 11001.  
XX  
KW Biochip; gene expression; gut; diagnostic; detection; probe; ss.  
XX  
OS Escherichia coli.  
XX  
PN EPI260592-A1.  
XX  
PD 27-NOV-2002.  
XX  
PF 17-MAY-2001; 2001EP-00112179.  
XX  
PR 17-MAY-2001; 2001EP-00112179.  
XX  
PA (MMGB-) MMG-BIOTECH AG.  
XX  
PI Donner H, Drescher B, Huber A, Weber J;  
XX  
DR WPI; 2003-241155/24.  
XX

PT Biochip containing probes complementary with open reading frames in  
PT Escherichia coli K12, useful for detecting gene expression and expression  
PT patterns.  
XX  
PS Claim 3; Page 1714; 2004pp; German.  
XX  
CC This invention describes a novel biochip comprising probe spots, each  
CC containing many identical probes. The probes are nucleotide sequences of  
CC 30-80 bases, are prepared ex situ from synthetic oligonucleotides and at  
CC least one includes a segment of at least 20 bases identical with, or  
CC complementary to, a segment of an open reading frame (orf) of Escherichia  
CC coli K12. The biochip is used for specific detection of gene expression  
CC in K12 and for determining the gene expression pattern, e.g. for  
CC diagnostic determination of which E. coli strains are present in the gut,  
CC and to determine the effects of e.g. growth media on gene expression. The  
CC biochip provides as comprehensive as possible detection of the K12  
CC genome, with simultaneous analysis of many different genes with a single  
CC device, and comparison of gene expression between K12 and its mutants or  
CC other E. coli strains in a single experiment. Apart from qualitative and  
CC quantitative information about gene expression, it also allows  
CC measurements of population densities for the various strains. The use of  
CC synthetic oligonucleotides for preparation of probes allows free  
CC variation in probe length and ensures high purity (and thus selectivity,  
CC reactivity and reproducibility); also synthetic probes are generally  
CC shorter than probes prepared by polymerase chain reaction. ACD68731 to  
CC ACD81540 represent oligonucleotide probes used with the biochip described  
CC in the invention  
XX  
SQ Sequence 100 BP; 16 A; 22 C; 26 G; 36 T; 0 U; 0 Other;  
  
Query Match 13.3%; Score 100; DB 8; Length 100;  
Best Local Similarity 100.0%; Pred. No. 1.1e-23;  
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 598 GGGCATCTGTTTACAGCGATTGTTTCCGCAATAGTGCAGTTAATATCGTCTTGC 657  
|||||  
Db 1 GGGCATCTGTTTACAGCGATTGTTTCCGCAATAGTGCAGTTAATATCGTCTTGC 60  
  
QY 658 CCTGTGCTGGTCTATTGGCGGATACGCAATTTCTATTAA 697  
|||||  
Db 61 CCTGTGCTGGTCTATTGGCGGATACGCAATTTCTATTAA 100  
  
RESULT 8  
ACD79727  
ID ACD79727 standard; DNA; 100 BP.  
XX  
AC ACD79727;  
XX  
DT 19-SEP-2003 (first entry)  
XX  
DE E. coli K12 MGI655 biochip probe SEQ ID 11003.  
XX  
KW Biochip; gene expression; gut; diagnostic; detection; probe; ss.  
XX  
OS Escherichia coli.  
XX  
PN EPI260592-A1.  
XX  
PD 27-NOV-2002.  
XX  
PF 17-MAY-2001; 2001EP-00112179.  
XX  
PR 17-MAY-2001; 2001EP-00112179.  
XX  
PA (MMGB-) MMG-BIOTECH AG.  
XX  
PI Donner H, Drescher B, Huber A, Weber J;  
XX  
DR WPI; 2003-241155/24.  
XX  
PT Biochip containing probes complementary with open reading frames in  
PT Escherichia coli K12, useful for detecting gene expression and expression

PT patterns.  
XX  
PS Claim 3; Page 1714; 2004pp; German.  
XX  
CC This invention describes a novel biochip comprising probe spots, each  
CC containing many identical probes. The probes are nucleotide sequences of  
CC 30-80 bases, are prepared ex situ from synthetic oligonucleotides and at  
CC least one includes a segment of at least 20 bases identical with, or  
CC complementary to, a segment of an open reading frame (orf) of Escherichia  
CC coli K12. The biochip is used for specific detection of gene expression  
CC in K12 and for determining the gene expression pattern, e.g. for  
CC diagnostic determination of which E. coli strains are present in the gut,  
CC and to determine the effects of e.g. growth media on gene expression. The  
CC biochip provides as comprehensive as possible detection of the K12  
CC genome, with simultaneous analysis of many different genes with a single  
CC device, and comparison of gene expression between K12 and its mutants or  
CC other E. coli strains in a single experiment. Apart from qualitative and  
CC quantitative information about gene expression, it also allows  
CC measurements of population densities for the various strains. The use of  
CC synthetic oligonucleotides for preparation of probes allows free  
CC variation in probe length and ensures high purity (and thus selectivity,  
CC reactivity and reproducibility); also synthetic probes are generally  
CC shorter than probes prepared by polymerase chain reaction. ACD68731 to  
CC ACD81540 represent oligonucleotide probes used with the biochip described  
CC in the invention  
XX  
SQ Sequence 100 BP; 23 A; 22 C; 25 G; 30 T; 0 U; 0 Other;  
XX  
Query Match 13.3%; Score 100; DB 8; Length 100;  
Best Local Similarity 100.0%; Pred. No. 1.1e-23;  
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 401 GACGGACTTCAGGCAAAACCAATCAGCTTTGGGCCAAGCTTGTGCAAGTTGTGAAC 460  
Db 1 GACGGACTTCAGGCAAAACCAATCAGCTTTGGGCCAAGCTTGTGCAAGTTGTGAAC 60  
QY 461 GTCAAATCATTGTGTACGGTGTACGGCAGCTGTGACGT 500  
Db 61 GTCAAATCATTGTGTACGGTGTACGGCAGCTGTGACGT 100  
RESULT 9  
ABK73143  
ID ABK73143 standard; DNA; 573 BP.  
XX  
AC ABK73143;  
XX  
DT 13-AUG-2002 (first entry)  
XX  
DE Bacillus licheniformis genomic sequence tag (GST) #434.  
XX  
KW Differential gene expression; genomic sequenced tag; GST;  
KW altered culture condition; environmental stress;  
KW physiological provocation; ds.  
XX  
OS Bacillus licheniformis.  
XX  
PN WO200229113-A2.  
XX  
PD 11-APR-2002.  
XX  
PF 05-OCT-2001; 2001WO-US031437.  
XX  
PR 06-OCT-2000; 2000US-00680598.  
PR 27-MAR-2001; 2001US-0279526P.  
XX  
PA (NOVO ) NOVOZYMES BIOTECH INC.  
PA (NOVO ) NOVOZYMES AS.  
XX  
PI Berka R, Clausen IG;  
XX  
DR WPI; 2002-416684/44.  
XX

PT Monitoring differential expression of several genes in first Bacillus  
PT cell relative to expression of same genes in one or more second Bacillus  
PT cells, by using substrate containing Bacillus genomic sequenced tag  
PT array.  
XX  
PS Claim 4; SEQ ID NO 434; 200pp; English.  
XX  
CC The invention describes a method of monitoring differential expression of  
CC genes in a first Bacillus cell relative to expression of the genes in  
CC other Bacillus cells, comprising hybridising labelled nucleic acid probes  
CC isolated from Bacillus cells to a substrate containing array of Bacillus  
CC genomic sequenced tags (GST), examining the array, and determining  
CC relative gene expression by an observed hybridisation reporter signal of  
CC a spot in the array. The method is useful for measuring the expression of  
CC genes in a first Bacillus cell relative to expression of the same genes  
CC in one or more second Bacillus cells. The method is useful for monitoring  
CC global expression of several genes from a Bacillus cell, discovering new  
CC genes, identifying possible functions of unknown open reading frames and  
CC monitoring gene copy number variation and stability. Monitoring changes  
CC in expression of genes may be used to provide a representation of the way  
CC in which Bacillus cells adapt to changes in culture conditions,  
CC environmental stress or other physiological provocation. Extensive follow  
CC -up characterisation is unnecessary, when one spot on an array equals one  
CC gene or one open reading frame, since sequence information is available.  
CC This sequence represents a genomic sequence tag (GST) used in the method  
CC of the invention. Note: The sequence data for this patent did not form  
CC part of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 573 BP; 128 A; 116 C; 129 G; 200 T; 0 U; 0 Other;  
XX  
Query Match 6.7%; Score 50.4; DB 6; Length 573;  
Best Local Similarity 46.6%; Pred. No. 4.2e-06;  
Matches 194; Conservative 0; Mismatches 221; Indels 1; Gaps 1;  
QY 124 TTTAAGTCTTTTGGACTTACACCCCTGATACCGCTATGACGCCGCAACCAATAT 183  
Db 6 TATTATAGCTTTCTTCGTATGTGATTAATGACATCCATTACGCCGCCGCAACCAATAT 65  
QY 184 TCTGCCCTTACCTCTGCTACGTCGCATGGAATTCGTCAAAGTACCGCGTGTGCAGG 243  
Db 66 TTTAATGATGATGAAGCGCAAGAGTCCGCTTTACAGGTTTCATGGCGTTTACGACGG 125  
QY 244 GATGAGTCTGGATTTTGTATGTGATGTTACTGTGTGCGGCATTTTCATTTCACTGGC 303  
Db 126 TATCTTGGCGGGTTTGACAGTACTTGGGATTTCTACGCGGTGCCCTTACAAATCAGCCTGTA 185  
QY 304 AGTGATTGACCCCGCAGCGGTACACCTTTGAGTTGGCGGGGGCGGCATATATTGCTG 363  
Db 186 CAATTGATTCCTGTTGTAGAGCCTTATTTTAACTTGCCGCGCGCTTATTGATTTA 245  
QY 364 GCTGGCGTGAATAATCGCCACACGACCCCAAGAAAGAGACGACTTCAGGCAAAACCAAT 423  
Db 246 TTGGCTTTGCAAGTCGGCTTTACCAAAAATAAAAAACAGATTCCACAGAAGCCGCTC 305  
QY 424 CAGCTTTTGGCCAGCTTTGCTTTGCAAGTTTGAACGTCAAAAATCATTTGTACGGTGT 483  
Db 306 CTCTTTTATATCCGGCTTTATATTTACAGTATCAATATTAAGACATTTGTCTTCAT 365  
QY 484 T-ACGGCAGTGTCAAGCTTTGTTCTGCCGCAACACAGGCGTTAAGCTGGTAGTT 538  
Db 366 TAACTGTAATGAGCGGCTTGTGCTTGGCCGTTCAACCAATTCTTGAATCGACAATT 421  
RESULT 10  
AAS92926  
ID AAS92926 standard; cDNA; 2325 BP.  
XX  
AC AAS92926;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #28730.  
XX

XX	Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW	food supplement; medical imaging; diagnostic; genetic disorder; ss.
KW	
XX	
OS	Homo sapiens.
XX	
PN	WO200175067-A2.
XX	
PD	11-OCT-2001.
XX	
PF	30-MAR-2001; 2001WO-US008631.
XX	
PR	31-MAR-2000; 2000US-00540217.
PR	23-AUG-2000; 2000US-00649167.
XX	
PA	(HYSE-) HYSEQ INC.
XX	
PI	Dzmanac RT, Liu C, Tang YT;
XX	
DR	WPI; 2001-639362/73.
DR	P-PSDB; ABG28739.
XX	
PT	New isolated polynucleotide and encoded polypeptides, useful in
PT	diagnostics, forensics, gene mapping, identification of mutations
PT	responsible for genetic disorders or other traits and to assess
PT	biodiversity.

PS Claim 1; SEQ ID NO 28730; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences. (I) is useful as hybridisation probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used  
 CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (I) is useful in gene therapy techniques to restore normal  
 CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activity. The  
 CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic  
 CC coding sequences of the invention. Note: The sequence data for this  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 CC  
 SQ Sequence 2325 BP; 510 A; 570 C; 642 G; 603 T; 0 U; 0 Other;

Query Match	5.7%;	Score 42.6;	DB 5;	Length 2325;
Best Local Similarity	53.7%;	Pred. No. 0.0046;		
Matches 110;	Conservative 0;	Mismatches 94;	Indels 1;	Gaps 1;

Qy      465 AATCATTTTGTACGGTGTTACGGCACTGCAGCTTTGTTCTGCCGAACAAGCGCT 524  
         ||| ||| | | | | | | | | | | | | | | |  
Db    440 AAC TTTTTCCTCGCGGTAAGGCACATCCGTCGCCCAATTTC CAAATTA CTGGACT 499

QY 525 TAAGCTGGGTAGTTGGCGCTACGCTTTTGCTGGCGATGATTGGACGTTGGCAATGTGT 584

Db 500 GAAAACTCCTGCTCAAGCGCTTTTCCCTGGCGGATATTGTGTGACCTTTTGGAAAAATGTTT 559

Qy 585 GCTGGGCGCTGGCGGCATCTGTTCAGCGATTGTTTC-GCCAGTATGCTGCCAGTTA 643  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 560 CTTGGGGCCTGCGGGGATATTTTTAAACAATTTTTCTCCCAATAAGTCCCACATTA 619

Qy	644	AAATATCGTCTTGCCCTGTTGCTGG	668
Db	620	AAATCGGGGGTCCCCGTGTTGCTGG	644

RESULT 11  
AAS88917/c  
ID AAS88917 standard; cDNA; 3020 BP.

AC AAS88917;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #24721.

**KW** Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder; ss.

**OS Homo sapiens.**

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US008631.

PR 31-MAR-2000; 2000US-00540217.

PR 23-AUG-2000; 2000US-00649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

WPI; 2001-639362/73.

DR P-PSDB; ABG24730.

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.

PS Claim 1; SEQ ID NO 24721; 103bp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published/pct/sequences](http://ftp.wipo.int/pub/published/pct/sequences)

Sequence 3020 BP; 852 A; 763 C; 758 G; 647 T; 0 U; 0 Other;

Query Match	5.7%;	Score 42.6;	DB 5;	Length 3020;
Best Local Similarity	53.7%;	Pred. No. 0.0054;		
Matches 110; Conservative	0;	Mismatches 94;	Indels 1;	Gaps 1;

**OY**

465 AAATCATTTTGTACGGTGTTACGGCACTGTCGAGTTTGTTCTGCCGAACAACAGCGCT 524  
||| ||| | | | | | | | | | | | |  
**Db** 1888 AACTTTTTCCITTCGGGGTAAGGCACATCCGTGGCCAATTCTCAAATAA CTG GACT 1829

525 TAAGCTGGTAGTTGGCCGTACGCGTTTGCTGGCCATGATTGGGACGTTGGCAATGTGT 584



Db 1828 GAAAACTCCTGCTCAAGCGTTTTCCTCGCGATATTGTGACTTTTGAAAAATGTTT 1769  
Qy 585 GCTGGCGCGTGGCGGCATCTGTTTCAGCGATTGTTTC-GCCAGTATGGTCCGCACTTA 643  
Db 1768 CTTGGGGCCTGCGGGGGATATTTTAAAGATTTTCTCCCAATAAGTCCCCAATTA 1709  
Qy 644 AATATCGTCTTGCCCTGTGCTGG 668  
Db 1708 AAATCGGGGTTCCCTGTGCTGG 1684

RESULT 12

ADA31058  
ID ADA31058 standard; DNA; 669 BP.  
XX  
AC ADA31058;  
XX  
DT 20-NOV-2003 (first entry)  
XX  
DE DNA encoding Acinetobacter baumannii protein #2345.  
XX  
KM ds; gene; Acinetobacter baumannii; bacterial disease; antibacterial;  
KM vaccine; plant biocontrol agent.  
XX  
OS Acinetobacter baumannii.  
XX  
PN US6562958-B1.  
XX  
PD 13-MAY-2003.  
XX  
PF 04-JUN-1999; 99US-00328352.  
XX  
PR 09-JUN-1998; 98US-0088701P.  
XX  
PA (GENO-) GENOME THERAPEUTICS CORP.  
XX  
PI Breton G, Bush D;  
XX  
DR WPI; 2003-576092/54.  
DR P-PSDB; ADA35184.  
XX  
PT New Acinetobacter baumannii proteins and nucleic acids, useful as reagents  
PT for diagnosing a bacterial disease, as components of antibacterial  
PT vaccines, as targets for antibacterial drugs, or as biocontrol agents for  
PT plants.  
XX  
PS Example; SEQ ID NO 2345; 328bp; English.  
XX  
CC The invention relates to isolated Acinetobacter baumannii nucleic acids.  
CC The A. baumannii nucleic acids and polypeptides are useful as reagents  
CC for diagnosing a bacterial disease, as components of antibacterial  
CC vaccines, as targets for antibacterial drugs, to detect the presence of  
CC A. baumannii and other Acinetobacter species in a sample, in screening  
CC compounds for the ability to interfere with the A. baumannii life cycle  
CC or to inhibit A. baumannii infection, and as biocontrol agents for  
CC plants. The present sequence represents DNA encoding an A. baumannii  
CC protein.  
XX  
SQ Sequence 669 BP; 152 A; 133 C; 155 G; 229 T; 0 U; 0 Other;  
Query Match 5.6%; Score 42; DB 9; Length 669;  
Best Local Similarity 50.4%; Pred. No. 0.0036;  
Matches 133; Conservative 0; Mismatches 125; Indels 6; Gaps 1;

Qy 105 AACAACTGACACCGACCCCTTTAAGTGTCTTTGGACTTACACCCGTATACCGCTATGA 164  
Db 53 AGCATATGGCTTTTAAACATTTTATTGCAATTTTGGAGTGTCTCCATTCTTTTATTATTA 112  
Qy 165 CGCCAGGACCGAACAATATTCTGCCCCCTTAGCTCTGCTACGTCCGATGGATTTCGTCAAA 224  
Db 113 CGCCAGGGGCGAGACTGGGCTTACGCCATTTCGCGAGGTAATTAAAGGTAAAGT-----CG 166  
Qy 225 GTACCCGCGGTGCTGGCAGGATGAGTCTGGATTTTTGAATTGTGATGTACTGTGTCGGG 284

Db 167 TCGTCCCCCGCTGTGTCAGAGGTATGCTAATTCGGGCACCTTATTACGATTTTATGTAGCCG 226  
Qy 285 GCATTTCATTTTCACTGGCAGTGAATGACCCCGCAGCGGTACACTTTTGAGTGGGCGG 344  
Db 227 CTGCTGTGGTCTGCTTGTAGCAATAAACCAACTGCTCTGATGATTTCTTACTGTGCGG 286  
Qy 345 GGGCGGCATATATTGCTGCGCTGG 368  
Db 287 GTTCTGCCCTAATTATTATGATGG 310

RESULT 13

ADF02251  
ID ADF02251 standard; DNA; 636 BP.  
XX  
AC ADF02251;  
XX  
DT 12-FEB-2004 (first entry)  
XX  
DE Bacterial polynucleotide #2536.  
XX  
KM Proteus mirabilis infection; bacterial infection; antibacterial;  
KM immunostimulant; gene; ds.  
XX  
OS Proteus mirabilis.  
XX  
PN US6605709-B1.  
XX  
PD 12-AUG-2003.  
XX  
PF 05-APR-2000; 2000US-00543681.  
XX  
PR 09-APR-1999; 99US-0128706P.  
XX  
PA (GENO-) GENOME THERAPEUTICS CORP.  
XX  
PI Breton GL;  
XX  
DR WPI; 2003-895291/82.  
DR P-PSDB; ADF06423.  
XX  
PT New Proteus mirabilis polypeptides and polynucleotides, useful as  
PT reagents for diagnosis of bacterial disease, as components of  
PT antibacterial vaccines, as targets for antibacterial drugs, or as  
PT biocontrol agents for plants.  
XX  
PS Disclosure; SEQ ID NO 2536; 870pp; English.  
XX  
CC The invention relates to new Proteus mirabilis polypeptides and  
CC polynucleotides. The invention also relates to antibodies against the  
CC polypeptides, methods for producing the polypeptides, a method of  
CC generating vaccines for immunising an individual against P. mirabilis, a  
CC method for evaluating a compound for the ability to bind a P. mirabilis  
CC polypeptide and a method for screening test compounds for anti-bacterial  
CC activity. The polypeptides and polynucleotides are useful as molecular  
CC targets for diagnosing, preventing and treating pathological conditions  
CC resulting from bacterial infection, as reagents for diagnosis of  
CC bacterial diseases, as components of antibacterial vaccines, as targets  
CC for antibacterial drugs or as bio-control agents for plants. This  
CC sequence represents a Proteus mirabilis polynucleotide of the invention.  
XX  
SQ Sequence 636 BP; 158 A; 97 C; 144 G; 237 T; 0 U; 0 Other;  
Query Match 5.3%; Score 39.8; DB 10; Length 636;  
Best Local Similarity 46.5%; Pred. No. 0.02;  
Matches 128; Conservative 0; Mismatches 147; Indels 0; Gaps 0;

Qy 111 TGACACCGACCCCTTTTAAGTGTCTTTTGGACTTACACCCGTGATTACCGGTATGACGCCAG 170  
Db 32 TGACAAATAGCCTAGTGTCTTTCGTTGAGCGTTTCTTTTATTGACGCAATTACGCCCG 91  
Qy 171 GACCGAACAATATTCTGCCCTTAGCTCTGCTACGTGCGATGATTTCGTCAAGTACC 230

Db 92 GACCTAATAATTACTATTAACTCTTCGGGGGCTCATGTGGTTAAACGCTTTAA 151  
QY 231 GCGTCTGGCAGGAGTACTGCGGATTTTGATGTGATGTTACTGTGCGGCAATT 290  
Db 152 CGTGTATGGCAGGCATTATCGTGGAAATGCAATGTATTACTTAGCTCGCTTGGTA 211  
QY 291 CATTTCACTGGCAGTATGACCCGGCAGCGGTACACCTTTTGAGTTGGCGGCGG 350  
Db 212 TTGCACATTATTGATTATTATTCCTGCATTGCATATCGGTTAAAGATTATTGGTACGA 271  
QY 351 CATATATTGCTGCTGGCGTGGAAATCGCCACC 385  
Db 272 TTATCTTTGTTAGCTTGAAACAGCAACC 306

RESULT 14  
ABD10542  
ID ABD10542 standard; DNA; 750 BP.  
XX  
AC ABD10542;  
XX  
DT 29-JUL-2004 (first entry)  
XX  
DE Pseudomonas aeruginosa polynucleotide #9146.  
XX  
KM Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;  
KM antibacterial.  
XX  
OS Pseudomonas aeruginosa.  
XX  
PN US6551795-B1.  
XX  
PD 22-APR-2003.  
XX  
PF 18-FEB-1999; 99US-00252991.  
XX  
PR 18-FEB-1998; 98US-0074788P.  
PR 27-JUL-1998; 98US-0094190P.  
XX  
PA (GENO-) GENOME THERAPEUTICS CORP.  
XX  
PI Rubenfield MJ, Nolling J, Deloughery C, Bush D;  
XX  
DR WPI; 2003-615309/58.  
DR P-PSDB; ABO76971.  
XX  
PT Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,  
PT useful as molecular targets for diagnostics, prophylaxis and treatment of  
PT pathological conditions resulting from bacterial infection.  
XX  
PS Disclosure; SEQ ID NO 9146; 455pp; English.  
XX

CC The invention relates to Pseudomonas aeruginosa polypeptides and the  
CC polynucleotides encoding them. The sequences are useful in diagnosis and  
CC therapy of pathological conditions, as molecular targets for diagnostics,  
CC prophylaxis and treatment of pathological conditions resulting from a  
CC bacterial infection, for evaluating a compound, such as a polypeptide,  
CC for the ability to bind a P. aeruginosa nucleic acid, as components of  
CC effective antibacterial targets, as targets for antibacterial drugs,  
CC including anti-P. aeruginosa drugs, as templates for recombinant  
CC production of P. aeruginosa-derived peptides or polypeptides, as target  
CC components for diagnosis and/or treatment of P. aeruginosa-caused  
CC infection, and in detection of P. aeruginosa sequences or other sequences  
CC of Pseudomonas species using biochip technology. Sequences ABD01397-  
CC ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:  
CC The sequence data for this patent did not form part of the printed  
CC specification but was obtained in electronic format from USPTO at  
CC [seqdata.uspto.gov/sequence.html](http://seqdata.uspto.gov/sequence.html)  
XX

SQ Sequence 750 BP; 113 A; 232 C; 248 G; 157 T; 0 U; 0 Other;  
Query Match 5.3%; Score 39.6; DB 11; Length 750;

Best Local Similarity 47.2%; Pred. No. 0.026;  
Matches 120; Conservative 0; Mismatches 134; Indels 0; Gaps 0;  
QY 131 GCTTTTGGACTTACACCTGATTACCGCTATGACGCCAGACCGAACAATTTCTGCC 190  
Db 160 GCGTTCTCCGTGTCGCTTCCGTACACCTCCGTACACGCCCGGTCCGAACAACCATGTTG 219  
QY 191 CTTAGCTCTGCTACGTCGCATGATTTGTCGAAAGTACCCGCGTGTGGCAGGATGAGT 250  
Db 220 CTGGCCTCGGGGTCAACTTCGGTTTCGTCGCTCCATCCGACATCCTGGGATCAGT 279  
QY 251 CTGGGATTTTGAATGTGATGTACTGTGTGCGGGCATTTTCAATTTTCACTGCAATGATT 310  
Db 280 TGCGGCTTCTTCAATCATGTGATGGCGGTGGGCTTCGGCCTGGGCACGGTGTTCGAGGCC 339  
QY 311 GACCCGGCAGCGGTACACCTTTGAGTTGGCGGGGGCGGCATATATTGTCGCTGGCG 370  
Db 340 TATCCGCTGTGTACACCATCTGCGCTACGTGCGCGCGGCTTACCTGCTGTACCTGGCC 399  
QY 371 TGGAAATCGCCAC 384  
Db 400 TGAAGATCGCCAC 413

RESULT 15  
ADL08109  
ID ADL08109 standard; DNA; 247682 BP.  
XX  
AC ADL08109;  
XX  
DT 20-MAY-2004 (first entry)  
XX  
DE Human gene associated with low HDL-C AT3.  
XX  
KM Human; ds; SNP; single nucleotide polymorphism;  
KM high density lipoprotein-C; HDL-C; vascular disease; metabolic disease;  
KM coronary artery disease; gene.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FH variation replace(77082,A)  
FT /\*tag= a  
FT /standard\_name= "Single nucleotide polymorphism"  
XX  
PN US2004043389-A1.  
XX  
PD 04-MAR-2004.  
XX  
PF 04-SEP-2002; 2002US-00235192.  
XX  
PR 04-SEP-2002; 2002US-00235192.  
XX  
PA (VITI-) VITIVITY INC.  
XX  
PI McCarthy J;  
XX  
DR WPI; 2004-214170/20.  
XX

PT Determining whether a subject has, or is at risk of developing, an  
PT abnormally low high density lipoprotein-C (HDL-C) level comprises  
PT detecting an allelic variant of a polymorphic region from any of a set of  
PT 27 genes.  
XX  
PS Disclosure; SEQ ID NO 28; 37pp; English.  
XX  
CC The invention relates to determining whether a subject has, or is at risk  
CC of developing, an abnormally low high density lipoprotein-C (HDL-C) level  
CC comprises determining whether the subject has an allelic variant of a  
CC polymorphic region from any of 27 genes (alleles listed in Table 5 of the  
CC specification). Also included are determining whether a male subject has,  
CC or is at risk of developing, an abnormally low HDL-C level, comprising  
CC determining whether the male subject has an allelic variant of a

polymorphic region listed in Table 5 which is associated with abnormally low HDL-C levels in males, and determining whether a female subject has, or is at risk of developing, an abnormally low HDL-C level, comprising determining whether the female subject has an allelic variant of a polymorphic region listed in Table 5 which is associated with abnormally low HDL-C levels in females. The allelic variant in determining whether a subject has, or is at risk of developing, an abnormally low HDL-C level is APOA 1 CC, CD14 1 CT, COL5A2 1 GG, EDNRB 1 AG or AA, FABP3 1 CT, GBE1 1 AG or GG, LIPC 5 AA, MTHFR 1 CC, VWF 2 GG, or their complements. The allelic variant in determining whether a male subject has, or is at risk of developing, an abnormally low HDL-C level, LRP1 3 CC or CT, PAI2 4 GG, or PPARG 1 CG, or their complements. The allelic variants are also COL5A2 1 GG, CD14 1 CT or CC, and FABP3 1 CT, in combination, or their complements. The methods are useful for diagnosing (a predisposition to) or abnormally low levels of low high density lipoprotein-C (HDL-C) in a subject. The methods are useful in diagnosing (a predisposition to) or prognosticating diseases and disorders associated with abnormal lipid levels such as vascular and metabolic diseases, e.g., coronary artery disease. The present sequence is a human gene containing a SNP (single nucleotide polymorphism associated with low high density lipoprotein-C (HDL-C) levels.

XX  
SQ Sequence 247682 BP; 75816 A; 50170 C; 48727 G; 72669 T; 0 U; 300 Other;

Query Match 5.2%; Score 39; DB 12; Length 247682;  
Best Local Similarity 55.6%; Pred. No. 1.2;  
Matches 75; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 6 ATAAACCCCAACCTATCGAAATATCGAATCTAGAAATATAAAACATTCTTTTAA 65  
Db 168128 ATAAACCTATATCTATGAAATTAATTGAATTATTAATAAACCTTCTCACAAGAAA 168187  
QY 66 TGTTCGCTGTCGGGTACTGTCTACCAAAACAGAGATTAACAAGTGACACCGACCCCTT 125  
Db 168188 ATGCCAGTCCAGATGGTGAATTATCAAAAATGTAAAGAAATTAACACTAATCCTAT 168247  
QY 126 TAAGTGCTTTTGA 140  
Db 168248 GCAAAGTCTTTCAGA 168262

Search completed: September 10, 2005, 07:59:20  
Job time : 479 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2005, 01:31:52 ; Search time 166 Seconds  
(without alignments)  
454.327 Million cell updates/sec

Title: US-10-620-487-2  
Perfect score: 996  
Sequence: 1 VTPTLLSAFWTYTLITAMP.....RQINIVLALLLVYCAVRIFY 195

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_16Dec04:\*

1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	996	100.0	195	8	ADK66006	Adk66006 E coli Yf
2	993	99.7	195	3	AAB01788	Aab01788 Escherich
3	684	68.7	205	7	ABO6144	Ab06144 Klebsiell
4	505	50.7	344	4	ABG17792	Abg17792 Novel hum
5	272.5	27.4	249	7	ABO76971	Ab076971 Pseudomon
6	256	25.7	203	6	ABM70439	Abm70439 Photorhab
7	250.5	25.2	211	7	ABO63925	Ab063925 Klebsiell
8	232.5	23.3	212	7	ADF04482	Adf04482 Bacterial
9	228.5	22.9	211	7	ADF06423	Adf06423 Bacterial
10	191.5	19.2	238	6	ADA35835	Ada35835 Acinetoba
11	178	17.9	241	6	ADA35872	Ada35872 Acinetoba
12	173	17.4	271	7	ABO83709	Ab083709 Pseudomon
13	163.5	16.4	235	6	ADA36962	Ada36962 Acinetoba
14	162	16.3	214	7	ABO61549	Ab061549 Klebsiell
15	161.5	16.2	206	6	ABU32233	Abu32233 Pseudon e
16	158.5	15.9	206	7	ABO70000	Ab070000 Pseudomon
17	158.5	15.9	211	6	ABU39959	Abu39959 Protein e
18	157.5	15.8	190	4	ABG29208	Abg29208 Novel hum
19	157.5	15.8	206	6	ABU28287	Abu28287 Protein e
20	156.5	15.7	206	6	ABU50100	Abu50100 Protein e
21	156.5	15.7	206	6	ABU47858	Abu47858 Protein e
22	155.5	15.6	214	6	ADA34738	Ada34738 Acinetoba
23	155.5	15.6	237	7	ABO78429	Ab078429 Pseudomon
24	154	15.5	205	6	ABU44939	Abu44939 Protein e
25	154	15.5	210	6	ADA35521	Ada35521 Acinetoba

26	153.5	15.4	211	7	ABO72001	Ab072001 Pseudomon
27	153	15.4	214	7	ABO63276	Ab063276 Klebsiell
28	151.5	15.2	211	6	ABU19902	Abu19902 Protein e
29	149.5	15.0	250	7	ADF04202	Adf04202 Bacterial
30	148	14.9	211	7	ABO63091	Ab063091 Klebsiell
31	146.5	14.7	220	6	ADA33616	Ada33616 Acinetoba
32	146	14.7	222	6	ADA35184	Ada35184 Acinetoba
33	143.5	14.4	212	3	ABAB01787	Aab01787 Escherich
34	138.5	13.9	219	7	ABO61214	Ab061214 Klebsiell
35	137.5	13.8	205	6	ABU49040	Abu49040 Protein e
36	132	13.3	208	6	ADA33824	Ada33824 Acinetoba
37	129	13.0	201	6	ABM68999	Abm68999 Photorhab
38	128.5	12.9	228	7	ADF04569	Adf04569 Bacterial
39	128.5	12.9	240	6	ADA36250	Ada36250 Acinetoba
40	128	12.9	220	7	ADF04260	Adf04260 Bacterial
41	127	12.8	200	6	ADA33486	Ada33486 Acinetoba
42	126	12.7	205	3	AAY79298	Aay79298 E. coli R
43	126	12.7	205	3	AAY99597	Aay99597 E. coli L
44	125.5	12.6	214	6	ADA35408	Ada35408 Acinetoba
45	123.5	12.4	209	6	ABU38890	Abu38890 Protein e

ALIGNMENTS

RESULT 1  
ADK66006  
ID ADK66006 standard; protein; 195 AA.  
XX AC ADK66006;  
XX DT 06-MAY-2004 (first entry)  
XX DE E coli Yfik protein.  
XX KW Yfik; phosphoglycerate amino acid; amino acid production.  
XX OS Escherichia coli.  
XX PN EPI382684-A1.  
XX PD 21-JAN-2004.  
XX PF 10-JUL-2003; 2003EP-00015546.  
XX PR 19-JUL-2002; 2002DE-01032930.  
XX PA (CONE ) CONSORTIUM ELEKTROCHEM IND GMBH.  
XX PI Maier T;  
XX DR WPI; 2004-192974/19.  
XX DR N-PSDB; ADK66005.  
XX PT New microbial strain, useful for producing amino acids of the  
PT phosphoglycerate family, particularly serine and cysteine, has increased  
PT activity of the yfik gene product.  
XX PS Disclosure; Page 10-11; 16pp; German.  
XX CC The present invention relates to a microbial strain for fermentative  
CC product of amino acids of the phosphoglycerate family, or their  
CC derivatives, has higher activity of the product of the yfik gene, or its  
CC homologue, compared with its parent strain. The microbial strain is used  
CC to produce O- and N-acetyl-serine, L-cysteine and cystine. The present  
CC sequence is the E. coli Yfik protein.  
SQ Sequence 195 AA;  
Query Match 100.0%; Score 996; DB 8; Length 195;  
Best Local Similarity 100.0%; Pred. No. 1.4e-109;  
Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VTPTLLSAFWTYTLITAMTPGPNNILALSSATSHGFRQSTRVLAGMSLGFLIVMLLCAGI 60  
Db 1 VTPTLLSAFWTYTLITAMTPGPNNILALSSATSHGFRQSTRVLAGMSLGFLIVMLLCAGI 60  
QY 61 SFSLAVIDPAVAHLLSWAGAAYIVWLAWKIATSPTEKEDGLQAKPISFWASFALQFVNVKI 120  
Db 61 SFSLAVIDPAVAHLLSWAGAAYIVWLAWKIATSPTEKEDGLQAKPISFWASFALQFVNVKI 120  
QY 121 ILYGVTALSTFVLPTQALSWVGVSVLLAMIGTGNVCWALAGHLFQRLFRQYGRQLNI 180  
Db 121 ILYGVTALSTFVLPTQALSWVGVSVLLAMIGTGNVCWALAGHLFQRLFRQYGRQLNI 180  
QY 181 VLALLLVYCAVRIFY 195  
Db 181 VLALLLVYCAVRIFY 195

RESULT 2  
AAB01788  
ID AAB01788 standard; proteain; 195 AA.

AC AAB01788;  
XX  
DT 03-JAN-2001 (first entry)  
XX  
DE Escherichia coli YfiK amino acid excretion protein.  
XX  
KM E. coli; yfiK gene; amino acid production; excretion protein gene;  
KM amino acid excretion protein.  
XX  
OS Escherichia coli.  
XX  
PN EP1016710-A2.  
XX  
PD 05-JUL-2000.  
XX  
PF 17-DEC-1999; 99EP-00125263.  
XX  
PR 30-DEC-1998; 98RU-00124016.  
PR 09-MAR-1999; 99RU-00104431.  
XX  
PA (AJIN ) AJINOMOTO CO INC.  
XX  
PI Livshits VA, Zakataeva NP, Nakanishi K, Aleshin VV, Troshin PV;  
PI Tokmakova IL;  
XX  
DR WPI; 2000-414802/36.  
DR N-PSDB; AAA52690.  
XX  
PT Increased production of L-amino acids by an Escherichia bacterium  
PT comprises increasing the expression amount of an L-amino acid excretion  
PT protein.  
XX  
PS Claim 1; Page 23; 29pp; English.  
XX  
CC The present sequence is the YfiK amino acid excretion protein from  
CC Escherichia coli. This protein is involved in the production of amino  
CC acids, and an increase in its expression leads to an increased  
CC accumulation of amino acids in the cell. In this case, an increase in  
CC threonine, histidine, glutamic acid and proline is achieved if multiple  
CC copies of its gene are transfected into a bacterium. The bacterium used  
CC is E. coli  
XX  
SQ Sequence 195 AA;

Query Match 99.7%; Score 993; DB 3; Length 195;  
Best Local Similarity 99.5%; Pred. No. 3.1e-109;  
Matches 194; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VTPTLLSAFWTYTLITAMTPGPNNILALSSATSHGFRQSTRVLAGMSLGFLIVMLLCAGI 60  
Db 1 VTPTLLSAFWTYTLITAMTPGPNNILALSSATSHGFRQSTRVLAGMSLGFLIVMLLCAGI 60

QY 61 SFSLAVIDPAVAHLLSWAGAAYIVWLAWKIATSPTEKEDGLQAKPISFWASFALQFVNVKI 120  
Db 61 SFSLAVIDPAVAHLLSWAGAAYIVWLAWKIATSPTEKEDGLQAKPISFWASFALQFVNVKI 120  
QY 121 ILYGVTALSTFVLPTQALSWVGVSVLLAMIGTGNVCWALAGHLFQRLFRQYGRQLNI 180  
Db 121 ILYGVTALSTFVLPTQALSWVGVSVLLAMIGTGNVCWALAGHLFQRLFRQYGRQLNI 180  
QY 181 VLALLLVYCAVRIFY 195  
Db 181 VLALLLVYCAVRIFY 195

RESULT 3  
ABO66144  
ID ABO66144 standard; proteain; 205 AA.

AC ABO66144;  
XX  
DT 29-JUL-2004 (first entry)  
XX  
DE Klebsiella pneumoniae polypeptide seqid 12661.  
XX  
KM Recombinant expression vector; transcription regulatory element;  
KM Klebsiella pneumoniae protein; antibacterial; Vaccine.  
XX  
OS Klebsiella pneumoniae.  
XX  
PN US6610836-B1.  
XX  
PD 26-AUG-2003.  
XX  
PF 27-JAN-2000; 2000US-00489039.  
XX  
PR 29-JAN-1999; 99US-0117747P.  
XX  
PA (GENO-) GENOME THERAPEUTICS CORP.  
XX  
PI Breton GL, Osborne M;  
XX  
DR WPI; 2003-895346/82.  
DR N-PSDB; ACH99695.  
XX  
PT New nucleic acid encoding a Klebsiella pneumoniae polypeptide, useful for  
PT preparing a vaccine composition against Klebsiella pneumoniae.  
XX  
PS Disclosure; SEQ ID NO 12661; 932pp; English.  
XX  
CC The invention describes a new isolated nucleic acid encoding a Klebsiella  
CC pneumoniae polypeptide. Also described are: a recombinant expression  
CC vector comprising the nucleic acid, operably linked to a transcription  
CC regulatory element; and a cell comprising the recombinant expression  
CC vector. The nucleic acid is useful for preparing a vaccine composition  
CC against Klebsiella pneumoniae. This is the amino acid sequence of a  
CC Klebsiella pneumoniae polypeptide of the invention  
XX  
SQ Sequence 205 AA;

Query Match 68.7%; Score 684; DB 7; Length 205;  
Best Local Similarity 68.4%; Pred. No. 1.8e-72;  
Matches 132; Conservative 21; Mismatches 40; Indels 0; Gaps 0;

QY 1 VTPTLLSAFWTYTLITAMTPGPNNILALSSATSHGFRQSTRVLAGMSLGFLIVMLLCAGI 60  
Db 10 VTPTLLSAFWTYTLITAMTPGPNNILALSSATSHGFRQSTRVLAGMSLGFLIVMLLCAGI 69  
QY 61 SFSLAVIDPAVAHLLSWAGAAYIVWLAWKIATSPTEKEDGLQAKPISFWASFALQFVNVKI 120  
Db 70 TFSLEVELDSRFTVLGWIGAYIILWLAQIAKSPATGTPSVEPVGFWASLGLOFVNVKI 129  
QY 121 ILYGVTALSTFVLPTQALSWVGVSVLLAMIGTGNVCWALAGHLFQRLFRQYGRQLNI 180  
Db 130 ILYGVTALSTFVLPTREPVMILISVLLAIGALGNLCWALAGHLFQRLFLLYGRQLNW 189

QY 181 VLALLLVYCAVRI 193  
Db 190 MLALLLVYCAVRI 202

RESULT 4

ABG17792  
ID ABG17792 standard; protein; 344 AA.  
AC ABG17792;  
XX 18-FEB-2002 (first entry)  
DT  
XX Novel human diagnostic protein #17783.  
DE  
XX  
KM Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX WO200175067-A2.  
PN  
XX 11-OCT-2001.  
PD  
XX 30-MAR-2001; 2001WO-US008631.  
PF  
XX 31-MAR-2000; 2000US-00540217.  
PR 23-AUG-2000; 2000US-00649167.  
XX  
XX (HYSE-) HYSEQ INC.  
PA  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
XX WPI; 2001-639362/73.  
DR N-PSDB; AAS81979.  
XX  
XX

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.

PS Claim 20; SEQ ID NO 48151; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
CC amino acid sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 344 AA;

Query Match 50.7%; Score 505; DB 4; Length 344;  
Best Local Similarity 57.6%; Pred. No. 6.7e-51;  
Matches 114; Conservative 14; Mismatches 40; Indels 30; Gaps 3;

QY 23 NNILASSATSHGFRQ-STRVLAGMSLGLIVMLCAGISFSLAVIDPA--AVHL----- 74

Db 29 NGMAVAVAKKLGMAPRQLAEQVLTHTLDLNGIASKVEIAGPGFINIFLDPAFLAEHVQALA 88  
QY 75 -----LSWAGAAIVWLAWKIATSPTEKEDGLQAKPISFWASFA 112  
Db 89 SDRLGVAATPEKQTIWVDYSAPNRGFFGGGAPLGTTLQDLTRPTKEDGLQAKPISFWASFA 148  
QY 113 LQFVNKIIIXGTALSTFVLPQTQALSWVGVSVLLAMIGTFGNVCMAAGHLFORLFR 172  
Db 149 LQFVNKIIIXGTALSTFVLPQTQALSWVGVSVLLAMIGTFGNVCMAAGHLFORLFR 208  
QY 173 QYGRQINIVLALLVYCA 190  
Db 209 QYGRQINIVLALLVYCA 226

RESULT 5

ABO76971  
ID ABO76971 standard; protein; 249 AA.  
AC ABO76971;  
XX 29-JUL-2004 (first entry)  
DT  
XX Pseudomonas aeruginosa polypeptide #9146.  
DE  
XX Bacterial infection; Pseudomonas aeruginosa infection; antibacterial.  
XX Pseudomonas aeruginosa.  
OS  
XX US6551795-B1.  
PN  
XX 22-APR-2003.  
PD  
XX 18-FEB-1999; 99US-00252991.  
PF  
XX 18-FEB-1998; 98US-0074788P.  
PR 27-JUL-1998; 98US-0094190P.  
XX  
XX (GENO-) GENOME THERAPEUTICS CORP.  
PA  
XX  
PI Rubenfield MJ, Nolling J, Deloughery C, Bush D;  
XX  
XX WPI; 2003-615309/58.  
DR N-PSDB; ABD10542.  
XX

PT Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,  
PT useful as molecular targets for diagnostics, prophylaxis and treatment of  
PT pathological conditions resulting from bacterial infection.

PS Disclosure; SEQ ID NO 25717; 455pp; English.

XX The invention relates to Pseudomonas aeruginosa polypeptides and the  
CC polynucleotides encoding them. The sequences are useful in diagnosis and  
CC therapy of pathological conditions, as molecular targets for diagnostics,  
CC prophylaxis and treatment of pathological conditions resulting from a  
CC bacterial infection, for evaluating a compound, such as a polypeptide,  
CC for the ability to bind a P. aeruginosa nucleic acid, as components of  
CC effective antibacterial targets, as targets for antibacterial drugs,  
CC including anti-P. aeruginosa drugs, as templates for recombinant  
CC production of P. aeruginosa-derived peptides or polypeptides, as target  
CC components for diagnosis and/or treatment of P. aeruginosa-caused  
CC infection, and in detection of P. aeruginosa sequences or other sequences  
CC of Pseudomonas species using biochip technology. Sequences ABO67826-  
CC ABO84396 represent P. aeruginosa polypeptides of the invention. Note: The  
CC sequence data for this patent did not form part of the printed  
CC specification but was obtained in electronic format from USPTO at  
CC seqdata.uspto.gov/sequence.html  
XX  
SQ Sequence 249 AA;

Query Match 27.4%; Score 272.5; DB 7; Length 249;  
Best Local Similarity 37.0%; Pred. No. 1.8e-23;



Matches	71;	Conservative	39;	Mismatches	67;	Indels	15;	Gaps	8
OY	5	LLSAFWTYTLITAMTPGPNNTIALISSATSHGFRÖSTRVLAGMSLGLIIVMLLCAGISFSL	64						
Db	51	LLLAFLSLFAFVTSVTPGPNNTMLLASGVNFGFVRSIPHILGISCGFFIMVM--AVGFGL	107						
OY	65	AVI--DPAAVHLISWAGAAIVLWAKIATS-PTKEDGL-QAKPISFWASFALQFVNK	119						
Db	108	GTVEAYPVLTYTLRYVGAAYLLIWLAKIATSGPAGNDPEGKPLSYWGAAAFQWVNP	167						
OY	120	IILYGVTLSTFVLPTQTALSWVGVSVLLAMIG--TFGNVCWALAGHLFQRLFR--QYG	175						
Db	168	AWMAVAGALSTYT-PLQGYFTNVVVISVTEFALINAPTIG--IWAQGSMLRNVL	224						
OY	176	RQLNIVLALLIV	187						
Db	225	RVFNVMALLIV	236						
RESULT 6									
ABM70439	ID	ABM70439	standard; protein; 203 AA.						
XX	AC	ABM70439;							
DT	20-NOV-2003	(first entry)							
DE	Photorhabdus luminescens	protein sequence #3536.							
XX	XX								
KW	Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;								
KW	detection; food; gene expression; plant; animal; microorganism; toxin;								
KW	antibiotic; biopesticide; virulence factor; disease model; plague;								
KW	whooping cough.								
XX	XX								
OS	Photorhabdus luminescens.								
XX	XX								
PN	WO200294867-A2.								
PD	28-NOV-2002.								
XX	XX								
PF	07-FEB-2002; 2002WO-IB003040.								
XX	XX								
PR	07-FEB-2001; 2001FR-00001659.								
XX	XX								
PA	(INSP ) INST PASTEUR.								
XX	(CNRS ) CNRS CENT NAT RECH SCI.								
PI	Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;								
XX	Buchrieser C;								
DR	WPI; 2003-148459/14.								
PT	Genomic sequence of Photorhabdus luminescens and encoded polypeptides,								
XX	useful e.g. as therapeutic antimicrobials and agricultural pesticides.								
PS	Claim 2; SEQ ID NO 3536; 1205pp; French.								
XX	XX								
CC	The invention relates to the isolation of genes and their encoded								
CC	proteins from Photorhabdus luminescens. The isolated sequences are								
CC	sources of probes and primers for detecting the genome of P. luminescens								
CC	and related species; to study polymorphisms; for gene analysis and for								
CC	detection/amplification of the genes. Antibodies (Ab) raised against the								
CC	polypeptides encoded by the genes are used for detection/identification								
CC	of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that								
CC	carry a gene-containing vector are used to select compounds that								
CC	modulate, regulate, induce or inhibit expression of the genes in plants,								
CC	animals or microorganisms other than P. luminescens and are able to alter								
CC	response or sensitivity to toxins and antibiotics produced by P.								
CC	luminescens. Cells transformed to express the genes are useful for								
CC	recombinant production of the proteins, particularly toxins and								
CC	antibacterials useful as insecticides, bactericides and fungicides. The								
CC	genes, proteins, vectors containing the genes and Ab are also useful								
CC	therapeutically (to treat microbial infection by bacteria or fungi that								

[illegible]

RESULT 7  
 ABO63925  
 ID ABO63925 standard; protein; 211 AA.  
 XX  
 AC ABO63925;  
 XX  
 DT 29-JUL-2004 (first entry)  
 XX  
 DE Klebsiella pneumoniae polypeptide seqid 10442.  
 XX  
 KW Recombinant expression vector; transcription regulatory element;  
 KW Klebsiella pneumoniae protein; antibacterial; Vaccine.  
 XX  
 OS Klebsiella pneumoniae.  
 XX  
 PN US6610836-B1.  
 XX  
 PD 26-AUG-2003.  
 XX  
 PF 27-JAN-2000; 2000US-00489039.  
 XX  
 PR 29-JAN-1999; 99US-0117747P.  
 XX  
 PA (GENO-) GENOME THERAPEUTICS CORP.  
 XX  
 PI Breton GL, Osborne M;  
 DR WPI; 2003-895346/82.  
 DR N-PSDB; ACH97476.  
 XX  
 PT New nucleic acid encoding a Klebsiella pneumoniae polypeptide, useful for  
 PT preparing a vaccine composition against Klebsiella pneumoniae.  
 XX  
 PS Disclosure; SEQ ID NO 10442; 932pp; English.  
 XX  
 CC The invention describes a new isolated nucleic acid encoding a Klebsiella  
 CC pneumoniae polypeptide. Also described are: a recombinant expression  
 CC vector comprising the nucleic acid, operably linked to a transcription  
 CC regulatory element; and a cell comprising the recombinant expression  
 CC vector. The nucleic acid is useful for preparing a vaccine composition  
 CC against Klebsiella pneumoniae. This is the amino acid sequence of a  
 CC Klebsiella pneumoniae polypeptide of the invention  
 XX  
 SQ Sequence 211 AA;

	Query Match	25.2%	Score 250.5;	DB 7;	Length 211;
	Best Local Similarity	32.4%;	Pred. No. 6e-21;		
	Matches	61;	Conservative	42;	Mismatches 78; Indels 7; Gaps 5;
QY	5	LLSAFTYTLITAMTPGPNNILALSSATSHGFRQSTRVLAGMSLGLIWMLLCAGISFSL	64		
Db	17	ILIAFLIYAFVTSITPGPNNTMLLASGLNYGFKRS�PHYLGISIGFAL-MVIAVGTGLGR	75		
QY	65	AVIDPAVHL-LSWAGAAYIWLMWKIATS--PTKEDGIQAKPISFWASFALQFVNVKII	121		
Db	76	VFENMPQVYIALRICGATYLIYLAWNIAATALPMISKISESARPFSEWBAAGFQWNPKAW	135		
QY	122	LYGVTAISTFVLPTQTALSINVGVSVLLAMIGTFGNVCWALAGHLFQRLFR--QYGRQLN	179		
Db	136	IMATGAITYT-LPQGENVRSVIFLSILPASINAPSVSIWVTFGATLRHWLNDIKYLRIFN	194		
QY	180	IYALALLV	187		
Db	195	VCMAMLLL	202		

RESULT 8  
 ID ADF04482 standard; protein; 212 AA.  
 AC ADF04482;  
 DT 12-FEB-2004 (first entry)  
 DE Bacterial polypeptide #595.  
 KW Proteus mirabilis infection; bacterial infection; antibacterial;  
 KW immunostimulant.  
 OS Proteus mirabilis.  
 PN US6605709-B1.  
 PD 12-AUG-2003.  
 PF 05-APR-2000; 2000US-00543681.  
 PR 09-APR-1999; 99US-0128706P.  
 PA (GENO-) GENOME THERAPEUTICS CORP.  
 PI Breton GL;  
 DR WPI; 2003-895291/82.  
 DR N-PSDB; ADF00310.  
 PT New Proteus mirabilis polypeptides and polynucleotides, useful as  
 PT reagents for diagnosis of bacterial disease, as components of  
 PT antibacterial vaccines, as targets for antibacterial drugs, or as  
 PT biocontrol agents for plants.  
 PS  
 PS Disclosure; SEQ ID NO 4767; 870pp; English.  
 CC The invention relates to new Proteus mirabilis polypeptides and  
 CC polynucleotides. The invention also relates to antibodies against the  
 CC polypeptides, methods for producing the polypeptides, a method of  
 CC generating vaccines for immunising an individual against P. mirabilis, a  
 CC method for evaluating a compound for the ability to bind a P. mirabilis  
 CC polypeptide and a method for screening test compounds for anti-bacterial  
 CC activity. The polypeptides and polynucleotides are useful as molecular  
 CC targets for diagnosing, preventing and treating pathological conditions  
 CC resulting from bacterial infection, as reagents for diagnosis of  
 CC bacterial diseases, as components of antibacterial vaccines, as targets  
 CC for antibacterial drugs or as bio-control agents for plants. This  
 CC sequence represents a Proteus mirabilis polypeptide of the invention.  
 SQ Sequence 212 AA;

[illegible]

RESULT 9  
 ADF06423 ID ADF06423 standard; protein; 211 AA.  
 XX  
 AC ADF06423;  
 XX  
 DT 12-FEB-2004 (first entry)  
 XX  
 DE Bacterial polypeptide #2536.  
 XX  
 KW Proteus mirabilis infection; bacterial infection; antibacterial;  
 immunostimulant.  
 XX  
 OS Proteus mirabilis.  
 XX  
 PN US6605709-B1.  
 XX  
 PD 12-AUG-2003.  
 XX  
 PF 05-APR-2000; 2000US-00543681.  
 XX  
 PR 09-APR-1999; 99US-0128706P.  
 XX  
 PA (GENO-) GENOME THERAPEUTICS CORP.  
 XX  
 PI Breton GI;  
 XX  
 DR WPI: 2003-895291/82.  
 DR N-PSDB; ADF02251.  
 XX  
 PT New Proteus mirabilis polypeptides and polynucleotides, useful as  
 PT reagents for diagnosis of bacterial disease, as components of  
 PT antibacterial vaccines, as targets for antibacterial drugs, ~~or~~ as  
 PT biocontrol agents for plants.  
 XX  
 PS Disclosure; SEQ ID NO 6708; 870pp; English.  
 XX  
 XX The invention relates to new Proteus mirabilis polypeptides and  
 polynucleotides. The invention also relates to antibodies against the  
 polypeptides, methods for producing the polypeptides, a method of  
 generating vaccines for immunising an individual against P. mirabilis, a  
 method for evaluating a compound for the ability to bind a P. mirabilis  
 polypeptide and a method for screening test compounds for anti-bacterial  
 activity. The polypeptides and polynucleotides are useful as molecular  
 targets for diagnosing, preventing and treating pathological conditions  
 resulting from bacterial infection, as reagents for diagnosis of  
 bacterial diseases, as components of antibacterial vaccines, as targets  
 for antibacterial drugs or as bio-control agents for plants. This  
 sequence represents a Proteus mirabilis polypeptide of the invention.  
 XX  
 SD Sequence 211 AA;

Query Match	22.9%;	Score 228.5;	DB 7,	length 211;
Best local Similarity	31.5%;	Pred. No. 2.5e-18;		
Matches	62;	Conservative 43;	Mismatches 81;	Indels 11; Gaps 7;
QY	1 VTPTLLSAFWTTYTLITAMTPGPNNILASSATSHGFROSTRVLAGMSLGLIVMLLCA-G	59		
	: :::	::   ::	:	:: ::
Db	11 MTNSLVSLSVFLFIATTEPGNNLLTSSGAHVGLKRSLTIMAGIIVGMQCVELLSAFC	70		
QY	60 ISFSLAVIDPAVHL-LSWAGAAYIVWLAWKIATSPRTKEGDAQ--PISFWASFALQF	115		
	:::	:: ::	:	:::
Db	71 IA-TLLIYP-ALHIGLKIGSIYLCWLAWKTATASYQRLDISSKVQTVAFOGGLQF	128		
QY	116 VNVKIILYGVTALSTFEVLPQTQALSWMVGVSVLLAMIGTFGNVCWALAGHLFORLFRQYG	175		
	:: ::	:	:::::	:
Db	129 LNPKAMMMGLGAVGSFSLAGDGYLGISGVISVMLLVNFIAGWNWILLGGTFISR-FLQSR	187		
QY	176 RQ---LNIVLALLLVYC	189		
	:::	:		
Db	188 RAMPLENIIMGILTAMC	204		

RESULT 10	
ADA35835	
ID	ADA35835 standard; protein; 238 AA.
XX	
AC	ADA35835;
XX	
DT	20-NOV-2003 (first entry)
XX	
DE	Acinetobacter baumannii protein #2996.
XX	
KW	Acinetobacter baumannii; bacterial disease; antibacterial; vaccine;
KW	plant biocontrol agent.
XX	
OS	Acinetobacter baumannii.
XX	
PN	US6562958-B1.
XX	
PD	13-MAY-2003.
XX	
PF	04-JUN-1999; 99US-00328352.
XX	
PR	09-JUN-1998; 98US-0088701P.
XX	
PA	(GENO-) GENOME THERAPEUTICS CORP.
XX	
PI	Breton G, Bush D;
XX	
DR	WPI; 2003-576092/54.
DR	N-PSDB; ADA31709.
XX	
PT	New Acinetobacter baumannii proteins and nucleic acids, useful as reagents
PT	for diagnosing a bacterial disease, as components of antibacterial
PT	vaccines, as targets for antibacterial drugs, or as biocontrol agents for
PT	plants.
XX	
PS	Example; SEQ ID NO 7122; 328bp; English.
XX	
CC	The invention relates to isolated Acinetobacter baumannii nucleic acids.
CC	The A. baumannii nucleic acids and polypeptides are useful as reagents
CC	for diagnosing a bacterial disease, as components of antibacterial
CC	vaccines, as targets for antibacterial drugs, to detect the presence of
CC	A. baumannii and other Acinetobacter species in a sample, in screening
CC	compounds for the ability to interfere with the A. baumannii life cycle
CC	or to inhibit A. baumannii infection, and as biocontrol agents for
CC	plants. The present sequence represents the amino acid sequence of an A.
CC	baumannii protein.
XX	
SQ	Sequence 238 AA;
Query Match	19.2%; Score 191.5; DB 6; Length 238;
Best Local Similarity	27.1%; Pred. No. 7.3e-14;

	Matches	52;	Conservative	46;	Mismatches	79;	Indels	15;	Gaps	8;
QY	6	LSAFWTYTLTAMTPGPNNILALSSATSHGRQSTRVLAGMSLGFLIVMLCAGISFSLA	65							
Db	41	LTAFLSLFAIVASITPGPTNFILSLSSHYKISKTLPVILGSCIGALLVY-VGIGLGST	99							
QY	66	VID-PAAVHLLSWAGAAIYIVLAWKIAF-SP--TKEDGLQAKPISFWASFALQFVNVKLI	121							
Db	100	ILAYPVIQKIMAMSGLIWLTVLAWKLNYNNEVISLEKNEQYPPIGFKAFIMQAINPKTW	159							
QY	122	LYGVLTALSTFVLPQTQALSWVGVSVLLAMIGTFGNVC--WALAGHLFORLFRQYGRQL	178							
Db	160	MMAFAVISVYT---KQGQDILVNVSSILSCIFLLIAFPCLYLMALVGRLSSTRLLSK-PKHI	215							
QY	179	NI---VLALLLV	187							
Db	216	NIENKIMAILLL	227							

```

RESULT 11
ID ADA35872 standard; protein; 241 AA.
XX
AC ADA35872;
XX
DT 20-NOV-2003 (first entry)
XX
DE Acinetobacter baumannii protein #3033.
XX
KW Acinetobacter baumannii; bacterial disease; antibacterial; vaccine;
XX plant biocontrol agent.
OS Acinetobacter baumannii.
XX
PN US6562958-B1.
XX
PD 13-MAY-2003.
XX
PF 04-JUN-1999; 99US-00328352.
XX
PR 09-JUN-1998; 98US-0088701P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Breton G, Bush D;
XX
XX WPI: 2003-576092/54.
DR N-PSDB; ADA31746.
XX
PT New Acinetobacter baumannii proteins and nucleic acids, useful as reagents
PT for diagnosing a bacterial disease, as components of antibacterial
PT vaccines, as targets for antibacterial drugs, or as biocontrol agents for
PT plants.
XX
PS Example; SEQ ID NO 7159; 328pp; English.
XX
XX The invention relates to isolated Acinetobacter baumannii nucleic acids.
CC The A. baumannii nucleic acids and polypeptides are useful as reagents
CC for diagnosing a bacterial disease, as components of antibacterial
CC vaccines, as targets for antibacterial drugs, to detect the presence of
CC A. baumannii and other Acinetobacter species in a sample, in screening
CC compounds for the ability to interfere with the A. baumannii life cycle
CC or to inhibit A. baumannii infection, and as biocontrol agents for
CC plants. The present sequence represents the amino acid sequence of an A.
CC baumannii protein.
XX
XX Sequence 241 AA;

Query Match 17.9%; Score 178; DB 6; Length 241;
Best Local Similarity 24.6%; Pred. No. 3e-12;
Matches 56; Conservative 48; Mismatches 78; Indels 46; Gaps 10;

1 VTPTLISAF-----WTYTL-----ITAMTPGPNIIALLSSATSHGPROSTRVL 43

```



Db 17 MNPTYLHFPISIRWTYMTAILIPYLIATLTLTTPGLDTTLIRRTATLEGKSKAFQAA 76  
QY 44 AGMSLGFLI--VMLCAGISFSLAVIDPAVHLLSWAGAAIYVWLAKIATSPTK----- 96  
Db 77 LGISLGCIAWGIIVAC-GLG-ALLMASDLAFNLLKWMGAAVLAWLGLNMLKKPRSOADI 134  
QY 97 EDGLQAKPIS---FWASPALQFVNKIIILYGVLTALSTFVLPTQALSMVWGVSVLLAMIG 153  
Db 135 QDNHSNRSTSENFIFKGFPGNLNPKVGIFYISFLPQFIIPAQASAVTWVWGLVMIHVIG 194  
QY 154 TEGNVCWA-----LAGHLFQRLFRQYGRQLN----IVLALLLVY 188  
Db 195 ----VLWSSLLILAMQPLSRYLKQPKFVKYMDRITGSIFVLFAKLAF 238

RESULT 12  
ABO83709  
ID ABO83709 standard; protein; 271 AA.

XX ABO83709;  
XX 29-JUL-2004 (first entry)  
XX Pseudomonas aeruginosa polypeptide #15884.  
XX Bacterial infection; Pseudomonas aeruginosa infection; antibacterial.  
XX Pseudomonas aeruginosa.  
XX US6551795-B1.  
XX 22-APR-2003.  
XX 18-FEB-1999; 99US-00252991.  
XX 18-FEB-1998; 98US-0074788P.  
PR 27-JUL-1998; 98US-0094190P.  
XX (GENO-) GENOME THERAPEUTICS CORP.  
XX Rubenfeld MJ, Nolling J, Deloughery C, Bush D;  
XX WPI; 2003-615309/58.  
DR N-PSDB; ABD17280.  
XX Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,  
PT useful as molecular targets for diagnostics, prophylaxis and treatment of  
PT pathological conditions resulting from bacterial infection.  
XX Disclosure; SEQ ID NO 32455; 455pp; English.

XX The invention relates to Pseudomonas aeruginosa polypeptides and the  
CC polynucleotides encoding them. The sequences are useful in diagnosis and  
CC therapy of pathological conditions, as molecular targets for diagnostics,  
CC prophylaxis and treatment of pathological conditions resulting from a  
CC bacterial infection, for evaluating a compound, such as a polypeptide,  
CC for the ability to bind a P. aeruginosa nucleic acid, as components of  
CC effective antibacterial targets, as targets for antibacterial drugs,  
CC including anti-P. aeruginosa drugs, as templates for recombinant  
CC production of P. aeruginosa-derived peptides or polypeptides, as target  
CC components for diagnosis and/or treatment of P. aeruginosa-caused  
CC infection, and in detection of P. aeruginosa sequences or other sequences  
CC of Pseudomonas species using biochip technology. Sequences ABO67826-  
CC ABO84396 represent P. aeruginosa polypeptides of the invention. Note: The  
CC sequence data for this patent did not form part of the printed  
CC specification but was obtained in electronic format from USPTO at  
CC seqdata.uspto.gov/sequence.html

XX Sequence 271 AA;  
Query Match 17.4%; Score 173; DB 7; Length 271;  
Best Local Similarity 27.9%; Pred. No. 1.4e-11;

Matches 56; Conservative 36; Mismatches 79; Indels 30; Gaps 7;  
QY 5 LLSAFWTTYLTITAMTPGPNNILALSSATSHGFRQSTRVLAGMSLGLIVMLCAGISFSL 64  
Db 73 LLLAMGAFSLSLSSISGPNVLTIVASGANHGFRRTLFPVTGATLGF-VLLIAFVGFWVR 131  
QY 65 AV-IDPAVHLLSNAGAAIYVWLAKIATSPTK---EDGLQAKPISFWASPALQFVNK 119  
Db 132 AIEAYPRFFDYLGMAAFAIAHVGRYRIATADPRLLAEENGVP---GFFQGVLLQWLNPK 187  
QY 120 IILYGVLTALSTFVLPTQALSMVWGVSVLLAMIGTFGNVC-----WALAGHLF----- 167  
Db 188 AWIACASGVALFASPSTHA-----PLLVFMAIYLVVCYLSLAAWALGDRVALLLDS 239  
QY 168 QRLFRQYGRQLNIVLALLLVY 188  
Db 240 PRRVRLFNRAWGTVLVTAGY 260

RESULT 13  
ADA36962  
ID ADA36962 standard; protein; 235 AA.

XX ADA36962;  
XX 20-NOV-2003 (first entry)  
XX Acinetobacter baumannii protein #4123.  
XX Acinetobacter baumannii; bacterial disease; antibacterial; vaccine;  
KW plant biocontrol agent.  
XX Acinetobacter baumannii.  
XX US6562958-B1.  
XX 13-MAY-2003.  
XX 04-JUN-1999; 99US-00328352.  
PF 09-JUN-1998; 98US-0088701P.  
PR (GENO-) GENOME THERAPEUTICS CORP.  
XX Breton G, Bush D;  
XX WPI; 2003-576092/54.  
DR N-PSDB; ADA32836.  
XX New Acinetobacter baumannii proteins and nucleic acids, useful as reagents  
PT for diagnosing a bacterial disease, as components of antibacterial  
PT vaccines, as targets for antibacterial drugs, or as biocontrol agents for  
PT plants.  
XX Example; SEQ ID NO 8249; 328pp; English.

XX The invention relates to isolated Acinetobacter baumannii nucleic acids.  
CC The A. baumannii nucleic acids and polypeptides are useful as reagents  
CC for diagnosing a bacterial disease, as components of antibacterial  
CC vaccines, as targets for antibacterial drugs, to detect the presence of  
CC A. baumannii and other Acinetobacter species in a sample, in screening  
CC compounds for the ability to interfere with the A. baumannii life cycle  
CC or to inhibit A. baumannii infection, and as biocontrol agents for  
CC plants. The present sequence represents the amino acid sequence of an A.  
CC baumannii protein.

XX Sequence 235 AA;  
Query Match 16.4%; Score 163.5; DB 6; Length 235;  
Best Local Similarity 31.4%; Pred. No. 1.5e-10;  
Matches 50; Conservative 23; Mismatches 73; Indels 13; Gaps 5;  
QY 8 AFWTTYLTITAMTPGPNNILALSSATSHGFRQSTRVLAGMSLGLIVMLCAGISF-SLAV 66

Db 34 AFGLICLAMVLTGPNMITYLISRSISQGIAGFISLGVAVGFVFM-LCASFGITALV 92  
QY 67 IDPAVHLISWAGAAIYVWLAKIATSPTKEDGLQAKPIS-----FWASFALQFNVK 119  
Db 93 AVPAYDTRIAGAMYLLMLWK-ALRPNAPIFNVKDLAVDSPKLFLMGFLTNLNPK 151  
QY 120 IILYGTALSTFVLPTQALSWVGVSVLLAMIGTEGNNV 158  
Db 152 IAIMYLSLPPQFIHPQGS--ILAQSIQLGTIQIFVSV 187

RESULT 14  
ABO61549  
ID ABO61549 standard; protein; 214 AA.  
XX  
AC ABO61549;  
XX  
DT 29-JUL-2004 (first entry)  
XX  
DE Klebsiella pneumoniae polypeptide seqid 8066.  
XX  
KM Recombinant expression vector; transcription regulatory element;  
KM Klebsiella pneumoniae protein; antibacterial; vaccine.  
XX  
OS Klebsiella pneumoniae.  
XX  
PN US6610836-B1.  
XX  
PD 26-AUG-2003.  
XX  
PF 27-JAN-2000; 2000US-00489039.  
XX  
PR 29-JAN-1999; 99US-0117747P.  
XX  
PA (GENO-) GENOME THERAPEUTICS CORP.  
XX  
PI Breton GL, Osborne M;  
XX  
DR WPI; 2003-895346/82.  
DR N-PSDB; ACH95100.  
XX  
XX  
PT New nucleic acid encoding a Klebsiella pneumoniae polypeptide, useful for  
PT preparing a vaccine composition against Klebsiella pneumoniae.  
XX  
PS Disclosure; SEQ ID NO 8066; 932pp; English.  
XX  
CC The invention describes a new isolated nucleic acid encoding a Klebsiella  
CC pneumoniae polypeptide. Also described are: a recombinant expression  
CC vector comprising the nucleic acid, operably linked to a transcription  
CC regulatory element; and a cell comprising the recombinant expression  
CC vector. The nucleic acid is useful for preparing a vaccine composition  
CC against Klebsiella pneumoniae. This is the amino acid sequence of a  
CC Klebsiella pneumoniae polypeptide of the invention  
XX  
SQ Sequence 214 AA;

Query Match 16.3%; Score 162; DB 7; Length 214;  
Best Local Similarity 26.5%; Pred. No. 2e-10;  
Matches 54; Conservative 42; Mismatches 82; Indels 26; Gaps 10;

QY 5 LISAFWTV-----TLITAMTPGPNNILALSSATSHGFRQSTRVLGMSLGFLLVMLCA- 58  
Db 8 LMTIEWWFAVLLTSTIILSPGSGAINTMTTSINHGYRGAAASIAGLQTLAIHIVLVGV 67  
QY 59 --GISFSLAVIDPAVHLISWAGAAIYVWLA--WKIATSPTKED--GLQAKPISFWASF 111  
Db 68 GLGTLFSRSVL--AFEVLKWAAGAAVLIWLGICQWRAGAIDLNTLAKAQTRGKLFQRAV 124  
QY 112 ALQFVNWKIILYGTALSTFVLPTQALSWV-VGSV---LLAMIG--TFGN--VCWA 161  
Db 125 FVNLTNPKSIVFLAALFPQFILPHQPOVMQYIVLGVTTIVVDIIVMIGYATLAQRISAWI 184

QY 162 LAGHLFQRLFRQYGRQLNIVTALL 185  
Db 185 KGPQKQKALNKVFGSLFMLVGALL 208

RESULT 15  
ABU32233  
ID ABU32233 standard; protein; 206 AA.  
XX  
AC ABU32233;  
XX  
DT 19-JUN-2003 (first entry)  
XX  
DE Protein encoded by Prokaryotic essential gene #17760.  
XX  
KM Antisense; prokaryotic essential gene; cell proliferation; drug design.  
XX  
OS Klebsiella pneumoniae.  
XX  
PN WO200277183-A2.  
XX  
PD 03-OCT-2002.  
XX  
PF 21-MAR-2002; 2002WO-US009107.  
XX  
PR 21-MAR-2001; 2001US-00815242.  
PR 06-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.  
XX  
PA (ELIT-) ELITRA PHARM INC.  
XX  
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX  
DR WPI; 2003-029926/02.  
DR N-PSDB; ACA36103.  
XX  
XX

PT New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX  
PS Claim 25; SEQ ID NO 60157; 1766pp; English.

XX  
CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway;  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than S. aureus, S. typhimurium,  
CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this



Page Blank (uspto)



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 10, 2005, 06:45:53 ; Search time 3392 Seconds  
(without alignments)  
10713.860 Million cell updates/sec

Title: US-10-620-487-1  
Perfect score: 750  
Sequence: 1 gatccataaccaccaacta.....ttagtaactgtctacttaag 750

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vi:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	DB	ID	Description
1	750	100.0	750	6	CQ759947	CQ759947 Sequence
2	750	100.0	10272	1	ECORBAB	D13169 E.coli gene
3	750	100.0	16446	1	D90886	D90886 E.coli geno
4	750	100.0	29254	1	ECOK12RITI	D64044 Escherichia
5	750	100.0	110000	1	U00096_27	Continuation (28 o
6	745.2	99.4	12322	1	AE015279	AE015279 Shigella
7	745.2	99.4	290380	1	AE016987	AE016987 Shigella
8	738.8	98.5	300099	1	AE016764	AE016764 Escherich
9	687.6	91.7	12337	1	AE005488	AE005488 Escherich
10	687.6	91.7	270365	1	AP002562	AP002562 Escherich
11	623.4	83.1	194171	2	AC126452	AC126452 Mus muscu
12	588	78.4	588	6	E49391	E49391 Process for
13	588	78.4	588	6	AX030083	AX030083 Sequence
14	461.2	61.5	145050	1	AL627275	AL627275 Salmonell
15	461.2	61.5	300431	1	AE016835	AE016835 Salmonell
16	458	61.1	21722	1	AE008820	AE008820 Salmonell
17	265.8	35.4	618	6	AR388761	AR388761 Sequence
18	224.4	29.9	23738	2	AC020828	AC020828 Mus muscu
19	156	20.8	372	6	AR388768	AR388768 Sequence

20	100	13.3	100	6	AX999538	AX999538 Sequence
21	100	13.3	100	6	AX999539	AX999539 Sequence
22	100	13.3	100	6	AX999540	AX999540 Sequence
23	72.6	9.7	110000	1	AE017333_05	Continuation (6 of
24	72.6	9.7	110000	1	CP000002_05	Continuation (6 of
25	70.4	9.4	288404	1	AE017272	AE017272 Bacillus
26	63.8	8.5	302451	1	AE016812	AE016812 Vibrio vu
27	55.2	7.4	110000	1	AE017225_22	Continuation (23 o
28	55.2	7.4	110000	1	AE017334_22	Continuation (23 o
29	55.2	7.4	110000	1	AE017355_23	Continuation (24 o
30	55.2	7.4	110000	1	CP000001_23	Continuation (24 o
31	55.2	7.4	293264	1	AE017031	AE017031 Bacillus
32	51	6.8	249500	1	AP005344	AP005344 Vibrio vu
33	50.4	6.7	573	6	AX432019	AX432019 Sequence
34	49.6	6.6	30600	1	AP005084	AP005084 Vibrio pa
35	49	6.5	110000	1	BX950851_45	Continuation (46 o
36	48.8	6.5	344805	1	BX640434_45	BX640434 Bordetell
37	48.8	6.5	346287	1	BX640450	BX640450 Bordetell
38	47.8	6.4	304262	1	AE017005	AE017005 Bacillus
39	47.2	6.3	13848	1	AE004249	AE004249 Vibrio ch
40	47.2	6.3	348171	1	BX640412	BX640412 Bordetell
41	44.4	5.9	310029	1	AE016868	AE016868 Pseudomon
42	42.8	5.7	142350	5	BX088596	BX088596 Zebrafish
43	42	5.6	669	6	AR319795	AR319795 Sequence
44	41.4	5.5	301214	1	AE016786	AE016786 Pseudomon
45	41	5.5	618	12	AY657639	AY657639 Synthetic

ALIGNMENTS

RESULT 1  
CQ759947  
LOCUS CQ759947 750 bp DNA linear PAT 03-MAR-2004  
DEFINITION Sequence 1 from Patent EP1382684.  
ACCESSION CQ759947  
VERSION CQ759947.1 GI:44903612  
KEYWORDS  
SOURCE  
ORGANISM Escherichia coli  
Escherichia coli  
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
Enterobacteriaceae; Escherichia.

REFERENCE  
AUTHORS Maier, T.  
TITLE process for fermentative production of amino acids and amino  
JOURNAL acid derivatives of the phosphoglycerate family  
Patent: EP 1382684-A 1 21-JAN-2004;  
Consortium fuer elektrochemische Industrie GmbH (DE)  
FEATURES  
source  
Location/Qualifiers  
1. 750  
/organism="Escherichia coli"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:562"  
110..697  
/note="unnamed protein product"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAF32791.1"  
/db\_xref="GI:44903613"  
/translation="MTPTLSAFWYTYLTITAMTPGPNNILASSATSHGFRQSTRVLA  
GMSLGLIVMLLCAGISFSLAVIDPAVHLLSWAGAYIVLWKIATSPTEKEDGLQA  
KPISEWASFALOFVNWKILYGVTAISTFVLPTQALSMVGVSVLLAMIGTFGNVCW  
ALAGHLFQRLFRQYGRQLNIVALLLVCAVRIFY"

ORIGIN

Query Match 100.0%; Score 750; DB 6; Length 750;  
Best local similarity 100.0%; Pred. No. 1.7e-224;  
Matches 750; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GATCCATAACCCCAAACTATCGAAATAATCGAATCTAGAAATATTAACATTCATTTT 60  
DB 1 GATCCATAACCCCAAACTATCGAAATAATCGAATCTAGAAATATTAACATTCATTTT 60

OY 61 TTAATGTTCCGTGTCGGGTA CTGTCTACCAAAA CAGAGAGATTA CAAGTGACACCGAC 120  
|||||  
Db 61 TTAATGTTCCGTGTCGGGTA CTGTCTACCAAAA CAGAGAGATTA CAAGTGACACCGAC 120  
OY 121 CTTTTAAGTCTTTTGGACTTACACCCGTGATTA CCGCTATGACGCCAGCAACAA 180  
|||||  
Db 121 CTTTTAAGTCTTTTGGACTTACACCCGTGATTA CCGCTATGACGCCAGCAACAA 180  
OY 181 TATTCGCCCCCTAGCTCTGCTAGCTGCATGGA TTTGTCAAAGTACCCGCGTGGC 240  
|||||  
Db 181 TATTCGCCCCCTAGCTCTGCTAGCTGCATGGA TTTGTCAAAGTACCCGCGTGGC 240  
OY 241 AGGATGAGTCTGGGATTTTGATGTGATGTGAT GTGTGCGGGCATTTCA TTTTCACT 300  
|||||  
Db 241 AGGATGAGTCTGGGATTTTGATGTGATGTGAT GTGTGCGGGCATTTCA TTTTCACT 300  
OY 301 GGCAGTATGACCCCGGACGGGTACACCTTTGAG TTGGGCGGGCGGCATATATTGT 360  
|||||  
Db 301 GGCAGTATGACCCCGGACGGGTACACCTTTGAG TTGGGCGGGCGGCATATATTGT 360  
OY 361 CTGGCTGGCGTGAATAATCGCCACCGACCCCA ACAAAGACGACTTCAGGCAAAACC 420  
|||||  
Db 361 CTGGCTGGCGTGAATAATCGCCACCGACCCCA ACAAAGACGACTTCAGGCAAAACC 420  
OY 421 AATCAGCTTTTGGGCGCACTTTGCTTGCAGTTT GCAAGCTCAAAATCATTTGTACGG 480  
|||||  
Db 421 AATCAGCTTTTGGGCGCACTTTGCTTGCAGTTT GCAAGCTCAAAATCATTTGTACGG 480  
OY 481 TGTACGGCACTGTGACGTTTGTCTGCGCAACA CAGGCGTTAAGCTGGGTAGTTGG 540  
|||||  
Db 481 TGTACGGCACTGTGACGTTTGTCTGCGCAACA CAGGCGTTAAGCTGGGTAGTTGG 540  
OY 541 CGTCAGCGTTTGTCTGGCGCATGATGGACGTTT GGCATGTGTCTGGCGCTGGCGGG 600  
|||||  
Db 541 CGTCAGCGTTTGTCTGGCGCATGATGGACGTTT GGCATGTGTCTGGCGCTGGCGGG 600  
OY 601 GCATCTGTTTACGCGATGTTTCCGACATATGTG TCCGCCAGTTAATATCGTGTGCCCT 660  
|||||  
Db 601 GCATCTGTTTACGCGATGTTTCCGACATATGTG TCCGCCAGTTAATATCGTGTGCCCT 660  
OY 661 GTTGCTGTTTATGCGCGGTACGCAATTTCTAT TAACGAAAAAAGCGGAAGGTGCG 720  
|||||  
Db 661 GTTGCTGTTTATGCGCGGTACGCAATTTCTAT TAACGAAAAAAGCGGAAGGTGCG 720  
OY 721 CCTCTCCGCTTAGTAATTGCTACTTAAG 750  
|||||  
Db 721 CCTCTCCGCTTAGTAATTGCTACTTAAG 750

RESULT 2  
ECORBAB 10272 bp DNA linear BCT 14-APR-2000  
LOCUS E.coli genes for L-aspartate oxidase, putative ATP dependent RNA  
DEFINITION helicase, and uracil DNA glycosylase.  
ACCESSION D13169  
VERSION D13169.1 GI:285775  
KEYWORDS ATP dependent RNA helicase; L-aspartate oxidase; uracil DNA  
glycosylase.  
SOURCE Escherichia coli  
ORGANISM Escherichia coli  
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
Enterobacteriaceae; Escherichia.  
REFERENCE 1 (bases 1 to 10272)  
AUTHORS Nashimoto,H., Miura,A., Saito,H. and Uchida,H.  
TITLE Suppressors of temperature-sensitive mutations in a ribosomal  
protein gene, rpsL (S12), of Escherichia coli K12  
JOURNAL Mol. Gen. Genet. 199 (3), 381-387 (1985)  
MEDLINE 85295454  
PUBMED 3162077  
REFERENCE 2 (sites)  
AUTHORS Varshney,U., Hutcheon,T. and van de Sande,J.H.  
TITLE Sequence analysis, expression, and conservation of Escherichia coli  
uracil DNA glycosylase and its gene (ung)

JOURNAL J. Biol. Chem. 263 (16), 7776-7784 (1988)  
MEDLINE 88227981  
PUBMED 2836397  
REFERENCE 3 (sites)  
AUTHORS Flachmann,R., Kunz,N., Seifert,J., Gutlich,M., Wientjes,F.J.,  
laufer,A. and Gassen,H.G.  
TITLE Molecular biology of pyridine nucleotide biosynthesis in  
Escherichia coli. Cloning and characterization of quinolinate  
synthesis genes nadA and nadB  
JOURNAL Eur. J. Biochem. 175 (2), 221-228 (1988)  
MEDLINE 88296484  
PUBMED 2841129  
REFERENCE 4 (sites)  
AUTHORS Nishi,K., Morel-Deville,F., Hershey,J.W., Leighton,T. and  
Schnier,J.  
TITLE An eIF-4A-like protein is a suppressor of an Escherichia coli  
mutant defective in 50S ribosomal subunit assembly  
JOURNAL Nature 336 (6198), 496-498 (1988)  
MEDLINE 89057150  
PUBMED 2461520  
REFERENCE 5 (sites)  
AUTHORS Nashimoto,H.  
TITLE Non-ribosomal proteins affecting the assembly of ribosomes in  
Escherichia coli  
JOURNAL (in) Nierhaus,K.H. (Ed.)  
THE TRANSLATIONAL APPARATUS: 185-195;  
Plenum Publishing Corp., New York, U.S.A. (1993)  
REFERENCE 6 (sites)  
AUTHORS Lonetto,M.A., Brown,K.L., Rudd,K.E. and Butner,M.J.  
TITLE Analysis of the Streptomyces coelicolor sigE gene reveals the  
existence of a subfamily of eubacterial RNA polymerase sigma  
factors involved in the regulation of extracytoplasmic functions  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 91 (16), 7573-7577 (1994)  
MEDLINE 94329558  
PUBMED 8052622  
REFERENCE 7 (bases 1 to 10272)  
AUTHORS Nashimoto,H.  
TITLE Direct Submission  
JOURNAL Submitted (02-SEP-1992) Hiroko Nashimoto, Teikyo University,  
Department of BioSciences, Toyosatodai, Utsunomiya, Tochigi 320,  
Japan (Tel:028-627-7181, Fax:028-627-7181)  
COMMENT Submitted (02-SEP-1992) to DDBJ by:  
Hiroko Nashimoto  
Dept. of Biosciences  
Teikyo University  
Toyosatodai, Utsunomiya  
Tochigi 320  
Japan  
Phone: 0286-27-7181  
Fax: 0286-27-7187.  
FEATURES  
source Location/Qualifiers  
1.10272  
/organism="Escherichia coli"  
/mol\_type="genomic DNA"  
/strain="K-12"  
/db\_xref="taxon:562"  
1642.3264  
/gene="nadB"  
1642.3264  
/gene="nadB"  
/citation=[3]  
/codon\_start=1  
/transl\_table=1  
/product="L-aspartate oxidase"  
/protein\_id="BAA02446.1"  
/db\_xref="GI:285776"  
/translation="MNTLPEHSCDVLIIISGAAGLSIALRLADQHOVIVLSKGPVTEG  
STFYAOGGIAAVFDETDSDSHVEDTLIAGAGICDRHAEFVASNARSCVQWLIDQV  
LPDTHIQNGEESYHLTREGGSHRILHAADATGREVEITLVSKALNPNIRVLERT  
NAVDLIVSDKIGLPGTRRVVGAWVNRNKEIVETCHAAVVLATGGASKVYQYTTNP  
ISSGDGIAMAMRAGCRVANLENQFHTALYHPQARNFLLEALRGEGAYLKRPDGTR  
FMPDFERGELAPRDIVARALDHEMKRLGADCMFLDISHKPADFIRQHPMIYEKLLG  
LGIDLTOEPVPIVPAHYTCGGVWVDHGRTDVEGLYAIGEVSYTGLHGANRMA NSL

LECLVYGSAAEDITRRMPYADISTLPWDESRVENPDERVIOHNMHELRLFMWDY  
VGIVRTKRLERLRITMLQGEIDERYAHFRVSNLLELRNLVQVAELIVRCAMMRK  
ESRGLHFTLDYPELLTHSGPSILSPGNHYINR"  
unsure 3094  
/gene="nadB"  
/citation=[3]  
/replace="g"  
3658  
/citation=[4]  
/replace="g"  
4117. .5451  
/gene="rbab"  
4117. .5451  
/gene="rbab"  
/citation=[5]  
/citation=[1]  
/citation=[4]  
/codon\_start=1  
/transl\_table=11  
/product="putative ATP dependent RNA helicase"  
/protein\_id="BAA02447.1"  
/db\_xref="GI:285777"  
/translation="MTVTTFSELESLDESLLEALQDKGFTRPATAOAAIPALDGRDV  
LGSAPGTGKTAAYLLPALQHLDFPRKSGPRLILPTRELAMQVSDHARELAKH  
THLDIATITGVAVMNABVFSENODIVATGRLLOYIKEENFDCRAVETLLIDEAD  
RMLDMGPAODIEHIAGETRWRKQTLFSATLEGDAIQDFAERLLEDPVEVSANPSTRE  
RKIHQWYRADDLLEHKTALLVHLKQPEATRSIVFVRKERVHELAWLREAGINNC  
YLEGEMVOGKRNEAIKRLTEGRVNVLVATDVAARGIDI PDVSHVFNPDMPRSGDTYLH  
RIGRTARAGRKGTALISLVEAHDDLGLKVGRYIEBPICARVIDELRPKTRAPSEKQTG  
KPSKVLARKRAEKKAKEKEKPRVKRHRDTKNIGKRKPSGTGVPQTTEE"  
variation 4531  
/gene="rbab"  
/note="HN1017"  
/citation=[5]  
/citation=[1]  
/phenotype="ribosome assembly defect, cold sensitive"  
/replace="c"  
4740  
/gene="rbab"  
/note="HN1016(double mutant)"  
/citation=[5]  
/citation=[1]  
/phenotype="ribosome assembly defect, cold sensitive"  
/replace="a"  
4966  
/gene="rbab"  
/note="HN1016(double mutant)"  
/citation=[5]  
/citation=[1]  
/phenotype="ribosome assembly defect, cold sensitive"  
/replace="a"  
variation 4966  
/gene="rbab"  
/note="HN1016(double mutant)"  
/citation=[5]  
/citation=[1]  
/phenotype="ribosome assembly defect, cold sensitive"  
/replace="c"  
5152  
/gene="rbab"  
/note="HN1007"  
/citation=[5]  
/citation=[1]  
/phenotype="ribosome assembly defect, cold sensitive"  
/replace="a"  
7979. .8668  
/gene="ung"  
7979. .8668  
/gene="ung"  
/citation=[2]  
/codon\_start=1  
/transl\_table=11  
/product="uracil DNA glycosylase"  
/protein\_id="BAA02448.1"  
/db\_xref="GI:285778"  
/translation="MANELTWHDLVLAEEKQPYFLNTLQTVASEROSGVTIYPQKDV  
FNAFRFTELGDVKVVIIGODPYHGPQAHGLAFSVRPGIAPPSLNMVKELENTIPG  
FTRPNHGYESWARQGVLLINTVLTVRAGQASHSHASIGWETFTDYISLINQHREGVV  
FLWGSHAQKGAIIDKQRHVLKAPRPSPLSAHRGFFGCNHPVLANQWLEQRGETPI  
DMMVLPABSE"

ORIGIN  
Query Match 100.0%; Score 750; DB 1; Length 10272;  
Best Local Similarity 100.0%; Pred. No. 3.1e-224;  
Matches 750; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GATCCATAACCCCAACCTATCGAAATATCGAATCTAGAATATATAAATTCATTTT 60  
DB 6539 GATCCATAACCCCAACCTATCGAAATATCGAATCTAGAATATATAAATTCATTTT 6598  
QY 61 TTAATGTTCCGTGTCGGGTACTGTCTACCAAAAACAGAGAGATACAAAGTGACACCGAC 120  
DB 6599 TTAATGTTCCGTGTCGGGTACTGTCTACCAAAAACAGAGAGATACAAAGTGACACCGAC 6658  
QY 121 CCTTTAAGTGTCTTTTGACTTACACCTGATTACCGCTATGACGCCAGAACAA 180  
DB 6659 CCTTTAAGTGTCTTTTGACTTACACCTGATTACCGCTATGACGCCAGAACAA 6718  
QY 181 TATTCGCCCTTAGCTCTGCTACGTCGCATGAGATTTCGTCAAAGTACCCGCTGCGC 240  
DB 6719 TATTCGCCCTTAGCTCTGCTACGTCGCATGAGATTTCGTCAAAGTACCCGCTGCGC 6778  
QY 241 AGGATGATGTCGGATTCTTTGATTGATGATGTTACTGTGTGCGGCAATTTCACTTTCAC 300  
DB 6779 AGGATGATGTCGGATTCTTTGATTGATGATGTTACTGTGTGCGGCAATTTCACTTTCAC 6838  
QY 301 GGCAGTATGACCCCGCAGCGGTACACCTTTGAGTTGGCGGGCGGCATATATTGT 360  
DB 6839 GGCAGTATGACCCCGCAGCGGTACACCTTTGAGTTGGCGGGCGGCATATATTGT 6898  
QY 361 CTGGCTGGCGTGAATAATCGCCACACGACCAACAAAGAGACGGACTTCAGGCCAAACC 420  
DB 6899 CTGGCTGGCGTGAATAATCGCCACACGACCAACAAAGAGACGGACTTCAGGCCAAACC 6958  
QY 421 AATCAGCTTTGGGCCAGCTTTGCTTGCAAGTTGTGAACGTCAAAATCATTTGTACCG 480  
DB 6959 AATCAGCTTTGGGCCAGCTTTGCTTGCAAGTTGTGAACGTCAAAATCATTTGTACCG 7018  
QY 481 TGTACGGCACTGTGCAAGTTGTTCTGCCCAACACAGAGCGTTAAGCTGGTAGTTGG 540  
DB 7019 TGTACGGCACTGTGCAAGTTGTTCTGCCCAACACAGAGCGTTAAGCTGGTAGTTGG 7078  
QY 541 CGTCAGCGTTTGTCTGGCGATGATGGGACGTTTGGCAATGTGTGCTGGCGCGCGG 600  
DB 7079 CGTCAGCGTTTGTCTGGCGATGATGGGACGTTTGGCAATGTGTGCTGGCGCGCGG 7138  
QY 601 GCATCTGTTTCAGCGATTGTTTCGCCAGTATGTCGCCAGTTAAATATCGTGTGCCCT 660  
DB 7139 GCATCTGTTTCAGCGATTGTTTCGCCAGTATGTCGCCAGTTAAATATCGTGTGCCCT 7198  
QY 661 GTTGCTGCTATTGCGCGGTACGCAATTTCTATTACGAAAAAAGCGGAAGAGTGGC 720  
DB 7199 GTTGCTGCTATTGCGCGGTACGCAATTTCTATTACGAAAAAAGCGGAAGAGTGGC 7258  
QY 721 CCTCTCCGCTTAGTAAGTTGCTACTTAAG 750  
DB 7259 CCTCTCCGCTTAGTAAGTTGCTACTTAAG 7288  
RESULT 3  
D90886 LOCUS D90886 16446 bp DNA linear BCT 29-MAY-1997  
DEFINITION E.coli genomic DNA, Kohara clone #436(58.4-58.8 min.).  
ACCESSION D90886 AB001340  
VERSION D90886.1 GI:1799980  
KEYWORDS Complete and shotgun sequencing; TRXA; kgtP; pss; pssA; ung; wltA;  
yfiD; yfiE; yfiF; yfiG; yfiK; yfiM.  
SOURCE Escherichia coli  
ORGANISM Escherichia coli  
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
Enterobacteriaceae; Escherichia.  
REFERENCE 1 (sites)  
AUTHORS Yamamoto,Y., Aiba,H., Baba,T., Hayashi,K., Inada,T., Isono,K.,



```

TITLE
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS
REFERENCE
97349980
9205837
2 (sites)
Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A., Horiuchi,T., Ikemoto,K., Inada,T., Isono,K., Isono,S., Itoh,T., Kanai,K., Kasai,H., Kashimoto,K., Kim,S., Kimura,S., Kitagawa,M., Kitakawa,M., Makino,K., Masuda,S., Miki,T., Mizobuchi,K., Mori,H., Motomura,K., Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T., Saito,N., Sampel,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C., Yamamoto,Y. and Yano,M.
The systematic sequencing of the Escherichia coli genome in Japan unpublished
3 (bases 1 to 16446)
Mori,H.
Direct Submission
Submitted (23-JAN-1997) Hirotsada Mori, NARA Institute of Science and Technology, Res. & Edu. Center for Genetic Info.; 8916-5 Takayama, Ikoma, Nara 630-01, Japan
(E-mail:hiori@gtc.aist-nara.ac.jp, Tel:81-7437-2-5660, Fax:81-7437-2-5669)
Collaboration Information:
Project:
The Japan E.coli genome DNA sequencing project
Group:
The Japan E.coli genome DNA sequencing group
Members: (1995.4 - 1996.3)
Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A., Horiuchi,T., Ikemoto,K., Inada,T., Isono,K., Isono,S., Itoh,T., Kanai,K., Kasai,H., Kashimoto,K., Kim,S., Kimura,S., Kitagawa,M., Kitakawa,M., Makino,K., Masuda,S., Miki,T., Mizobuchi,K., Mori,H., Motomura,K., Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T., Saito,N., Sampel,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C., Yamamoto,Y. and Yano,M.
Headed by:
Name: Takashi Horiuchi
Address: National Institute of Basic Biology, Okazaki, 444, Japan
E-mail: kishori@nibb.ac.jp
Information operator:
Name: Hirotsada Mori
Address: NARA Institute of Science and Technology, Ikoma, 630-01, Japan
E-mail: hmori@gtc.aist-nara.ac.jp
URL:
The Japan E. coli genome database
http:bsw3.aist-nara.ac.jp.
location/Qualifiers
1.16446
/organism="Escherichia coli"
/mol_type="genomic DNA"
/strain="K12"
/db_xref="taxon:562"
/map="58.4 min"
/clone_lib="Kohara clone #436"
/note="Nucleotide position 2709472-2725917 from the initiation site of ThrA (0 min).~This clone is from Kohara lambda minist library"
complement(1..623)
/gene="yfiB"
complement(<1..623)
/gene="yfiB"
/note="similar to [SwissProt Accession Number P33634]"
/codon_start=1

```

```

gene
CDS
    /transl_table=11
    /protein_id="BAA20920.1"
    /db_xref="GI:4433790"
    /translation="MDLRRFITLKTVEEGSFRLASQKLCTOSTVTFHIQQLQEPS
VQLFEKIGRRMCLTRREGKLLPHIYELTRVWDTLREAKKESDPDGLRVSGETLLS
YRMPQVQLQRFRRAPRKYRLSLQSLNCYIVRALINDEADVGFYRVGNDALNRRELG
EQLSLVASPQIADVDFTEPRGNACSFINEPQCVFRQIFESTLRQR"
    726..1313
    /gene="yfiK"
    /note="similar to [SwissProt Accession Number P38101]"
    /codon_start=1
    /transl_table=11
    /protein_id="BAA16464.1"
    /db_xref="GI:1799982"
    /translation="MTPTLLSAFWTYTLITAMTPGPNNILALSSATSHGRQSTRVLA
GMSLGLIWMLCAGISFSLAVIDPAVHLSWAGAAYIVLAWKIATSPKEDGLQA
KPLSFWASPALQFVNKILLYGTALSTFVLPTQALSWVGVSVLLAMIGTFGNVCW
ALAGHLFQRLFRQYGRQNLIVLALLLVCAVRIFY"
    complement(1369..1752)
    /gene="yfiD"
    /complement(1369..1752)
    /gene="yfiD"
    /note="similar to [SwissProt Accession Number P33633]"
    /codon_start=1
    /transl_table=11
    /protein_id="BAA16465.1"
    /db_xref="GI:1799983"
    /translation="MITGIQITKAANDLINSFWLLDSEKGEARCIYAKAGYAEDEVV
AVSLGDIIEYREVPEVKEPRVEGGQHLNVNLRLRETLDDAVKHPEKYPQLTIRVSG
YAVRNSLTPEQQRDVARTFTESL"
    2057..2746
    /gene="ung"
    /note="similar to [PIR Accession Number A28175]"
    /codon_start=1
    /transl_table=11
    /product="uracil-DNA glycosylase (EC 3.2.2.-)"
    /protein_id="BAA16466.1"
    /db_xref="GI:1799984"
    /translation="MANELTWDVDVLAEEKQRPFLNTLQTVASERQSGVTYPQKDV
FNAFRTELDVKVILIGQDYHGPQAHLGAFSVBGAIPSLINMYKELNTIPG
FLTRPHGYLESWARQGVLLNTVLTVRAGQASHASIGMETFTDKVISLINQHRGVV
FLMGSHAQKKGAIIDQRHNVLKAPHPSPLSAHRGFGCNHFLVANQWLQRGETPII
DMPVLPAAESE"
    complement(2577..3914)
    /gene="yfiF"
    /complement(2577..3914)
    /gene="yfiF"
    /note="similar to [SwissProt Accession Number P33635]"
    start codon is not identified yet"
    /codon_start=1
    /transl_table=11
    /protein_id="BAA16467.1"
    /db_xref="GI:1799985"
    /translation="SQALVRNHERPRGLRVICALRAALVEKSMNDEMKGSGKVVMY
VRSDSDSKRTNPRRTGKGGRGGRADGRRPARDDKQSPDRKWDSPWRVTSRS
APGDETEPKADHGIGSGKSFIDPEVLRQRAEETRVYGENACQALFQSRPEAIYRAHFA
IQSVTEPREKALRMMAANRKAHVVDDEALTKASGTEHGGVCFLIKRNQTTVOQWV
SQAGAQCVALALENESNPHNLGMMRSCAHFGVKGVVVQDALLLESAAIRTAEGGAB
HVQPTGDNIVNVLDLDFRQAGYTVVTTTSEQGPLFKTSIPAKMVLVLQGEYEGEDDA
ARDPNDLRVKIDGTGNVAGLNSVQPAFCLVNGGVRIKPKDSAGWQLCHPAFPRKFTTH
SLPVLASQSASRHHVPAITDLPAQNGCSQRIHDAFKATDAVLVSHDVAYL"
    4010..4243
    /gene="yfiG"
    /note="similar to [SwissProt Accession Number P33636]"
    start codon is not identified yet"
    /codon_start=1
    /transl_table=11

```



Query Match	100.0%; Score 750; DB 1; Length 16446;
Best Local Similarity	100.0%; Pred. No. 3.4e-224;
Matches	750; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GATCCATAACCCCAACCTATCGAAATATCGAATCTAGAATATAAACAATTCATTTT	60
Db 617 GATCCATAACCCCAACCTATCGAAATATCGAATCTAGAATATAAACAATTCATTTT	676
Qy 61 TTAATGTCGTCGCGGTACTGTCTACGAAACAGAGAGATTAACAAGTGACACCGAC	120
Db 677 TTAATGTCGTCGCGGTACTGTCTACGAAACAGAGAGATTAACAAGTGACACCGAC	736
Qy 121 CCTTTAAGTCCTTTTGGACCTTACACCCGTGATTAACGCTATGACGCCAGACCGAACA	180
Db 737 CCTTTAAGTCCTTTTGGACCTTACACCCGTGATTAACGCTATGACGCCAGACCGAACA	796
Qy 181 TATTCTGCCCTTAGCTCTGCTACGTCGCATGATTCGTCAAAGTACCCGCGTCTGCC	240
Db 797 TATTCTGCCCTTAGCTCTGCTACGTCGCATGATTCGTCAAAGTACCCGCGTCTGCC	856
Qy 241 AGGATGAGTCTGGGATTTTGTGATGTGATGTTACTGTGCGGCGCATTTTCACT	300
Db 857 AGGATGAGTCTGGGATTTTGTGATGTGATGTTACTGTGCGGCGCATTTTCACT	916
Qy 301 GGCAGTGAATGACCCGCGACGCGTACACCTTTTGAAGTGGCGGGCGGCATATATGT	360
Db 917 GGCAGTGAATGACCCGCGACGCGTACACCTTTTGAAGTGGCGGGCGGCATATATGT	976
Qy 361 CTGGCTGGCGTGAATAATGCCACGACGCCCAAGAAAGACGGAATTTCAGGCAAAACC	420

Db	977	CTGGCTGGCGGTGAAAAATCGCCACCCAGCCCAACAAAGGAAGACGGACTTCAGGCAAAACC	1036
Qy	421	AATCAGCTTTGGGCCAGCTTTGCTTTGCAGTTTGTAACGTCAAAATCATTTTGTACGG	480
Db	1037	AATCAGCTTTTGGGCCAGCTTTGCTTTGCAGTTTGTAACGTCAAAATCATTTTGTACGG	1096
Qy	481	TGTTACGGCACTGTGACGTTTGTCTGTCGCCGCAACACAGGCGTTAAGCTGGGTAGTTGG	540
Db	1097	TGTTACGGCACTGTGACGTTTGTCTGTCGCCGCAACACAGGCGTTAAGCTGGGTAGTTGG	1156
Qy	541	CGTCAGCGTTTGTGTCGGCATGATTGGGACGTTTGGCAATGTGTGCTGGCGCTGGCGGG	600
Db	1157	CGTCAGCGTTTGTGTCGGCATGATTGGGACGTTTGGCAATGTGTGCTGGCGCTGGCGGG	1216
Qy	601	GCACTGTGTTACGCCAGATTGTTTCCGCCAGTATGTCGCCAGTTAATATCGTCTTGCCCT	660
Db	1217	GCACTGTGTTACGCCAGATTGTTTCCGCCAGTATGTCGCCAGTTAATATCGTCTTGCCCT	1276
Qy	661	GTTGCTGCTCTATTGCGCGGTACGCATTTCATTAAACGAAAAAAGCGAAGAGTCCG	720
Db	1277	GTTGCTGCTCTATTGCGCGGTACGCATTTCATTAAACGAAAAAAGCGAAGAGTCCG	1336
Qy	721	CCTCTTCCGCTTAGTACTTGCTACTTAAG	750
Db	1337	CCTCTTCCGCTTAGTACTTGCTACTTAAG	1366
RESULT 4			
ECOK12RIII		29254 bp	DNA linear BCT 24-JAN-2003
LOCUS	ECOK12RIII	Escherichia coli genes for ribonuclease III and others, complete	
DEFINITION	Escherichia coli genes for ribonuclease III and others, complete		
ACCESSION	D64044		
VERSION	D64044.1	GI:987635	
KEYWORDS			
SOURCE	Escherichia coli		
ORGANISM	Escherichia coli		
	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;		
	Enterobacteriaceae; Escherichia.		
REFERENCE	1	Nashimoto, H. and Uchida, H.	
AUTHORS		DNA sequencing of the Escherichia coli ribonuclease III gene and	
TITLE		its mutations	
JOURNAL		Mol. Gen. Genet. 201 (1), 25-29 (1985)	
MEDLINE		86039802	
PUBMED		3903434	
REFERENCE	2	March, P.E. and Inouye, M.	
AUTHORS		Characterization of the lep operon of Escherichia coli.	
TITLE		Identification of the promoter and the gene upstream of the signal	
JOURNAL		peptidase I gene	
MEDLINE		J. Biol. Chem. 260 (12), 7206-7213 (1985)	
PUBMED		85207751	
REFERENCE	3	2987248	
AUTHORS		Ahnn, J., March, P.E., Takif, H.E. and Inouye, M.	
TITLE		A GTP-binding protein of Escherichia coli has homology to yeast RAS	
JOURNAL		proteins	
MEDLINE		Proc. Natl. Acad. Sci. U.S.A. 83 (23), 8849-8853 (1986)	
PUBMED		87067411	
REFERENCE	4	3097637	
AUTHORS		Varshney, U., Hutcheon, T. and van de Sande, J.H.	
TITLE		Sequence analysis, expression, and conservation of Escherichia coli	
JOURNAL		uracil DNA glycosylase and its gene (ung)	
MEDLINE		J. Biol. Chem. 263 (16), 7776-7784 (1988)	
PUBMED		88227981	
REFERENCE	5	2836397	
AUTHORS		Flachmann, R., Kunz, N., Seifert, J., Gutlich, M., Wientjes, F.J.,	
TITLE		Laufer, A. and Gassen, H.G.	
		Molecular biology of pyridine nucleotide biosynthesis in	
		Escherichia coli. Cloning and characterization of quinolinate	

synthesis genes *nadA* and *nadB*  
Eur. J. Biochem. 175 (2), 221-228 (1988)  
JOURNAL MEDLINE 88296484  
PUBMED 2841129  
REFERENCE  
AUTHORS  
TITLE  
6 Schendel,F.J., Mueller,E., Stubbe,J., Shiau,A. and Smith,J.M.  
Formylglycinamide ribonucleotide synthetase from *Escherichia coli*:  
cloning, sequencing, overproduction, isolation, and  
characterization  
Biochemistry 28 (6), 2459-2471 (1989)  
JOURNAL MEDLINE 89274173  
PUBMED 2659070  
REFERENCE  
AUTHORS  
TITLE  
7 Morrison,P.T., Lovett,S.T., Gilson,L.E. and Kolodner,R.  
Molecular analysis of the *Escherichia coli* *reco* gene  
J. Bacteriol. 171 (7), 3641-3649 (1989)  
JOURNAL MEDLINE 89291705  
PUBMED 2544549  
REFERENCE  
AUTHORS  
TITLE  
8 Poulsen,L.K., Larsen,N.W., Molin,S. and Andersson,P.  
Analysis of an *Escherichia coli* mutant strain resistant to the  
cell-killing function encoded by the *gef* gene family  
Mol. Microbiol. 6 (7), 895-905 (1992)  
JOURNAL MEDLINE 92292954  
PUBMED 1602968  
REFERENCE  
AUTHORS  
TITLE  
9 Nashimoto,H.  
Non-ribosomal proteins affecting the assembly of ribosomes in  
*Escherichia coli*  
(in) Nierhaus,K.H. (Ed.);  
THE TRANSLATIONAL APPARATUS: 185-195;  
Plenum Publishing Corp., New York, U.S.A. (1993)  
JOURNAL MEDLINE 95270615  
PUBMED 7751307  
REFERENCE  
AUTHORS  
TITLE  
10 Hiratsu,K., Amemura,M., Nashimoto,H., Shinagawa,H. and Makino,K.  
The *rpoE* gene of *Escherichia coli*, which encodes sigma E, is  
essential for bacterial growth at high temperature  
J. Bacteriol. 177 (10), 2918-2922 (1995)  
JOURNAL MEDLINE 95286510  
PUBMED 7768826  
REFERENCE  
AUTHORS  
TITLE  
11 Yu,H., Schurr,M.J. and Deretic,V.  
Functional equivalence of *Escherichia coli* sigma E and *Pseudomonas*  
*aeruginosa* AlgU: *E. coli* *rpoE* restores mucoidy and reduces  
sensitivity to reactive oxygen intermediates in algu mutants of *P.*  
*aeruginosa*  
J. Bacteriol. 177 (11), 3259-3268 (1995)  
JOURNAL MEDLINE 95286510  
PUBMED 7768826  
REFERENCE  
AUTHORS  
TITLE  
12 (bases 1 to 29254)  
Nashimoto,H. and Saito,N.  
Kohara library; 7G4-4A12 contig  
Unpublished  
13 (bases 1 to 29254)  
Nashimoto,H.  
Direct Submission  
Submitted (03-SEP-1995) Hiroko Nashimoto, Teikyo University,  
Department of Biosciences,; Toyosatodai, Utsunomiya, Tochigi 320,  
Japan (E-mail:nashimoto@nasu.bio.teikyo-u.ac.jp, Tel:81-28-627-7181,  
Fax:81-28-627-7181)  
FEATURES  
source  
1. .29254  
/organism="Escherichia coli"  
/mol\_type="genomic DNA"  
/strain="K12"  
/db\_xref="taxon:562"  
/clone\_lib="Kohara"  
/note="Clones 7G4, 4A12"  
3406. .4767  
/citation=[8]  
/codon\_start=1  
/transl\_table=11  
/product="YFED-ECOLI protein similarity"  
/protein\_id="BAA10908.1"

/db\_xref="GI:987636"  
/translation="MSTHTPLTYNEINGKPFGLDYELAKQFADYLGVKLKVTVRQNI  
SOLFDDLDNGNADLLAAGLYNSERVKNYQPGFTYYSVSCQLVYKVGQYRPTLGNLT  
ADRLTVA PGHVAVNDLQTLKETKPELSMKVDDKKGAELMEDVYBGKDYTLADSV  
ISLFORVHPELAVALDITDEQPTWESPLDGDNTLSAALLDFENEMEDGTARIEBK  
YLGHGDDPDYVDTRTFELRAVDVLPQLKPLFEKYAEEDWRLLAAIAYQESHWDQAT  
SPGVGRGMMMLTKNTAQSLGITDRDAEOSISGGVRYLQDMSKVPEVPENERIWF  
LAAVNMGYAHMLDARALTAKTGNPDSWADVQRPLPLISQPKYSKLTYYARGHEAY  
AYENIRKYQISLVGYLQEKKEQATEAMQLAQDYPAVSLRSWAKRNFLFSRFPSPRH  
QTI"  
3701. .3704  
/citation=[6]  
/citation=[8]  
/replace="c"  
3753. .3756  
/citation=[6]  
/citation=[8]  
/replace="g"  
3804. .3807  
/citation=[6]  
/citation=[8]  
/replace="g"  
4697. .4700  
/citation=[8]  
/replace="gcc"  
complement(4820. .5356)  
/citation=[8]  
/codon\_start=1  
/transl\_table=11  
/product="YFHC-ECOLI protein"  
/protein\_id="BAA10909.1"  
/db\_xref="GI:987637"  
/translation="MRRAFITGVFFLSEVEFSHEYNNRHALLTLAKRAWDEREVPVAV  
LVHNNRVIGEGWNRPIGRHDPYAAAEIMALROGLVQNYRLIDATLYVTLEPCWCA  
GAMHSRIGRVVFGARDAKTGAAGSLMDVLHPGMNHRVEITEGILADECAALLSDFF  
RMROEIKAQKKAQSSTD"  
complement(5381. .5995)  
/citation=[8]  
/codon\_start=1  
/transl\_table=11  
/product="YFHB-ECOLI protein"  
/protein\_id="BAA10910.1"  
/db\_xref="GI:987638"  
/translation="MVFFDLDGTLHQQDMFGSFLRYLRRQPLNALLVPLPITATA  
LILKGRARAWPMSLLWGCTFGHSEARLQLOADFVRWRDNTVAFPLVOERTTYLL  
SSDADIWLITGSPQPLVEAVYFDTPWLPVNLIASQIDRGYGGWVLTMRCLGHEKVAQ  
LERKIGTPLRLYSGYSDSNQNPLLYFCQHRWRVTPRGELQGLE"  
6153. .7073  
/citation=[8]  
/codon\_start=1  
/transl\_table=11  
/product="YFHH-ECOLI protein similarity"  
/protein\_id="BAA10911.1"  
/db\_xref="GI:987639"  
/translation="MEYKIHRSVSHIACCRHYFIARKDMNGLLRIRORYGLAQSDK  
LADVLLIQPDTARHLSSQQLANAGVSQSVVKFAQKLGKGPALKLALSEALASOP  
ESPSVPIHQIRGDDPLRLVGKLIKENTAAMYATLVNSEEKLIHECVTLRSARRIT  
LTGIGASGLVAQNFAWKLMKIGFNAAAVRDMHALATVQASSPDDLILAISYTGVRRE  
LNLADEMPLRVGKVLAITGFTFNALQQRAASHCLYTTIAEQATNSASISACHAQGMLT  
DLFTALIQDDLLELAPERIRHSEALVKLV"  
complement(9209. .9937)  
/citation=[7]  
/codon\_start=1  
/transl\_table=11  
/product="reco protein"  
/protein\_id="BAA10912.1"  
/db\_xref="GI:987640"  
/translation="MEGWQRAVFLHSRPMWSETSLMLDVTEESGRVRLVAKGARSKRS  
TLKGAQOPTPLLLRFGRGVKTLSAAVSLALPLSGITLYSGLYINELSRVLEY  
ETRFSELPFDYLHCIOSLAGVGTPEPALRRFELALLGLGYGVNFTHCAGSGEPVDD  
TMTYRYREEKGFIAVVIDNKTFTGRQLKALNAREFPDADTLRAAKRFTRMALKPYLG  
GKPLKSRLEFRQFMPKRTVKTXYE"  
complement(9949. .10854)  
gene

/gene="rbaA"  
complement(9949. .10854)  
/gene="rbaA"  
/citation={3}  
/citation={9}  
/codon\_start=1  
/transl\_table=11  
/product="GTP binding protein"  
/protein\_id="BAA10913.1"  
/db\_xref="GI:987641"

Query Match 100.0%; Score 750; DB 1; Length 29254;  
Best Local Similarity 100.0%; Pred. No. 3.9e-224;  
Matches 750; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATCCATAACCCCAACCTATCGAAAAATATCGAATCTAGAATATATAAACATTCAATTTT 60  
DB 22779 GATCCATAACCCCAACCTATCGAAAAATATCGAATCTAGAATATATAAACATTCAATTTT 22838  
QY 61 TTTAATGTTCCGTGTCGGGTACTGTCTACCAAAAACAGAGAGATAACAGTGACACCGAC 120  
DB 22839 TTTAATGTTCCGTGTCGGGTACTGTCTACCAAAAACAGAGAGATAACAGTGACACCGAC 22898  
QY 121 CCTTTAAGTCTTTTGGACTTACACCCCTGATTAACGCTATGACGCCAGACGAAACAA 180  
DB 22899 CCTTTAAGTCTTTTGGACTTACACCCCTGATTAACGCTATGACGCCAGACGAAACAA 22958  
QY 181 TATTCGCCCCCTAGCTCTGCTACGTCGATGATTCGTCAAAGTACCCGCGTGTGGC 240  
DB 22959 TATTCGCCCCCTAGCTCTGCTACGTCGATGATTCGTCAAAGTACCCGCGTGTGGC 23018  
QY 241 AGGATGAGTCTGGGATTTTGAATGTGATGTACTGTGCGGGCATTTCAATTTTCACT 300  
DB 23019 AGGATGAGTCTGGGATTTTGAATGTGATGTACTGTGCGGGCATTTCAATTTTCACT 23078  
QY 301 GGCAGTGATGACCCCGCAGCCGTACACCTTTTGGAGTTGGCGGGCGGCGCATATATTGT 360  
DB 23079 GGCAGTGATGACCCCGCAGCCGTACACCTTTTGGAGTTGGCGGGCGGCGCATATATTGT 23138  
QY 361 CTGGCTGGCGTGAATAATCGCCACACGACCCCAACAAAGAAAGACGACTTCAGGCAAAAAC 420  
DB 23139 CTGGCTGGCGTGAATAATCGCCACACGACCCCAACAAAGAAAGACGACTTCAGGCAAAAAC 23198  
QY 421 AATCAGCTTTTGGGCCACTTTGCTTTCAGAGTTTGAACGTCAAATCATTTTGTACGG 480  
DB 23199 AATCAGCTTTTGGGCCACTTTGCTTTCAGAGTTTGAACGTCAAATCATTTTGTACGG 23258  
QY 481 TGTACGGCACTGTGACGTTTGTCTGCCCGCAACACAGCGCTTAAGCTGGGTAGTTGG 540  
DB 23259 TGTACGGCACTGTGACGTTTGTCTGCCCGCAACACAGCGCTTAAGCTGGGTAGTTGG 23318  
QY 541 CGTCAGCGTTTGTGCGGCGATGATTGGAGCGTTTGGCAATGTGTGCTGGCGCGTGGCGGG 600  
DB 23319 CGTCAGCGTTTGTGCGGCGATGATTGGAGCGTTTGGCAATGTGTGCTGGCGCGTGGCGGG 23378  
QY 601 GCATCTGTTTACGCGATGTTTTCGCCAGTATGTCGCCAGTTAATATCGTGTGCCCT 660  
DB 23379 GCATCTGTTTACGCGATGTTTTCGCCAGTATGTCGCCAGTTAATATCGTGTGCCCT 23438  
QY 661 GTTGCTGCTATATGCGCGGTACGCAATTTTCTATTAAAGAAAAAGCGAAGAGTGC 720  
DB 23439 GTTGCTGCTATATGCGCGGTACGCAATTTTCTATTAAAGAAAAAGCGAAGAGTGC 23498  
QY 721 CCTCTCGCTTAAGTACTGCTACTTAAG 750  
DB 23499 CCTCTCGCTTAAGTACTGCTACTTAAG 23528

RESULT 5  
U00096.27  
WPCOMMENT  
Sequence split into 47 fragments LOCUS U00096 Accession U00096  
Fragment Name Begin End

U00096\_00 1 110000  
U00096\_01 100001 210000  
U00096\_02 200001 310000  
U00096\_03 300001 410000  
U00096\_04 400001 510000  
U00096\_05 500001 610000  
U00096\_06 600001 710000  
U00096\_07 700001 810000  
U00096\_08 800001 910000  
U00096\_09 900001 1010000  
U00096\_10 1000001 1110000  
U00096\_11 1100001 1210000  
U00096\_12 1200001 1310000  
U00096\_13 1300001 1410000  
U00096\_14 1400001 1510000  
U00096\_15 1500001 1610000  
U00096\_16 1600001 1710000  
U00096\_17 1700001 1810000  
U00096\_18 1800001 1910000  
U00096\_19 1900001 2010000  
U00096\_20 2000001 2110000  
U00096\_21 2100001 2210000  
U00096\_22 2200001 2310000  
U00096\_23 2300001 2410000  
U00096\_24 2400001 2510000  
U00096\_25 2500001 2610000  
U00096\_26 2600001 2710000  
U00096\_27 2700001 2810000  
U00096\_28 2800001 2910000  
U00096\_29 2900001 3010000  
U00096\_30 3000001 3110000  
U00096\_31 3100001 3210000  
U00096\_32 3200001 3310000  
U00096\_33 3300001 3410000  
U00096\_34 3400001 3510000  
U00096\_35 3500001 3610000  
U00096\_36 3600001 3710000  
U00096\_37 3700001 3810000  
U00096\_38 3800001 3910000  
U00096\_39 3900001 4010000  
U00096\_40 4000001 4110000  
U00096\_41 4100001 4210000  
U00096\_42 4200001 4310000  
U00096\_43 4300001 4410000  
U00096\_44 4400001 4510000  
U00096\_45 4500001 4610000  
U00096\_46 4600001 4639675  
Continuation (28 of 47) of U00096 from base 2700001 (U00096 Escherichia coli K-12 MG1655)

Query Match 100.0%; Score 750; DB 1; Length 110000;  
Best Local Similarity 100.0%; Pred. No. 5.2e-224;  
Matches 750; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATCCATAACCCCAACCTATCGAAAAATATCGAATCTAGAATATATAAACATTCAATTTT 60  
DB 13336 GATCCATAACCCCAACCTATCGAAAAATATCGAATCTAGAATATATAAACATTCAATTTT 13395  
QY 61 TTTAATGTTCCGTGTCGGGTACTGTCTACCAAAAACAGAGAGATTAACAAGTGACACCGAC 120  
DB 13396 TTTAATGTTCCGTGTCGGGTACTGTCTACCAAAAACAGAGAGATTAACAAGTGACACCGAC 13455  
QY 121 CCTTTAAGTCTTTTGGACTTACACCCCTGATTACCGCTATGACGCCAGACCGAAACAA 180  
DB 13456 CCTTTAAGTCTTTTGGACTTACACCCCTGATTACCGCTATGACGCCAGACCGAAACAA 13515  
QY 181 TATTCGCCCCCTAGCTCTGCTACGTCGATGATTCGTCAAAGTACCCGCGTGTGGC 240  
DB 13516 TATTCGCCCCCTAGCTCTGCTACGTCGATGATTCGTCAAAGTACCCGCGTGTGGC 13575  
QY 241 AGGATGAGTCTGGGATTTTGAATGTGATGTACTGTGTGCGGCGCATTTCAATTTTCACT 300  
DB 13576 AGGATGAGTCTGGGATTTTGAATGTGATGTACTGTGTGCGGCGCATTTCAATTTTCACT 13635



QY 301 GGACGTGATTGACCCGCGGTACACCTTTGAGTTGGCGGGCGGCATATATTGT 360  
|||||  
Db 13636 GGACGTGATTGACCCGCGCGGTACACCTTTGAGTTGGCGGGCGGCATATATTGT 13695  
QY 361 CTGGCTGGCGTGAANAATCGCCACCAGCCCAACAAGAGACGACTTCAGGCAAAACC 420  
|||||  
Db 13696 CTGGCTGGCGTGAANAATCGCCACCAGCCCAACAAGAGACGACTTCAGGCAAAACC 13755  
QY 421 AATCAGCTTTGGGCGCAGCTTTGCTTGCAGTTGTGAACGTCAAAATCATTTTGTACGG 480  
|||||  
Db 13756 AATCAGCTTTGGGCGCAGCTTTGCTTGCAGTTGTGAACGTCAAAATCATTTTGTACGG 13815  
QY 481 TGTTACGGCACTGTCGACGTTTGTCTGCCGCAACACAGGCGTTAAGCTGGGTAGTTGG 540  
|||||  
Db 13816 TGTTACGGCACTGTCGACGTTTGTCTGCCGCAACACAGGCGTTAAGCTGGGTAGTTGG 13875  
QY 541 CGTCACGCTTTTGTGCTGGCGATGATTGGACGTTTGGCAATGTGTGCTGGCGCTGGCGGG 600  
|||||  
Db 13876 CGTCACGCTTTTGTGCTGGCGATGATTGGACGTTTGGCAATGTGTGCTGGCGCTGGCGGG 13935  
QY 601 GCATCTGTTTACGCGATGTTTCCGACGATGTCGCCAGTTTAATATCGTCTTGCCCT 660  
|||||  
Db 13936 GCATCTGTTTACGCGATGTTTCCGACGATGTCGCCAGTTTAATATCGTCTTGCCCT 13995  
QY 661 GTTGCTGCTATATGCGCGGTACGCATTTTCTATTAAAGAAAAAAGCGAAGAGCTCGC 720  
|||||  
Db 13996 GTTGCTGCTATATGCGCGGTACGCATTTTCTATTAAAGAAAAAAGCGAAGAGCTCGC 14055  
QY 721 CCTCTTCGCTAGTACTTGCTACTTAAG 750  
|||||  
Db 14056 CCTCTTCGCTAGTACTTGCTACTTAAG 14085

RESULT 6  
AE015279 12322 bp DNA linear BCT 18-OCT-2002  
LOCUS Shigella flexneri 2a str. 301 section 242 of 412 of the complete genome.  
ACCESSION AE015279 AE005674  
VERSION AE015279.1 GI:24053018  
KEYWORDS  
SOURCE  
ORGANISM Shigella flexneri 2a str. 301  
Shigella flexneri 2a str. 301  
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Shigella.  
REFERENCE  
AUTHORS Jin,Q., Yuan,Z.H., Xu,J.G., Wang,Y., Shen,Y., Lu,W.C., Wang,J.H., Liu,H., Yang,J., Yang,F., Qu,D., Zhang,X.B., Zhang,J.Y., Yang,G.W., Wu,H.T., Dong,J., Sun,L.L., Xue,Y., Zhao,A.L., Gao,Y.S., Zhu,J.P., Kan,B., Chen,S.X., Yao,Z.J., He,B.K., Chen,R.S., Ma,D.L., Qiang,B.Q., Wen,Y.M., Hou,Y.D. and Yu,J.  
TITLE Genome sequence of Shigella flexneri 2a: insights into pathogenicity through comparison with genomes of Escherichia coli K12 and O157  
JOURNAL Nucleic Acids Res. 30 (20), 4432-4441 (2002)  
PUBMED 12384590  
REFERENCE  
AUTHORS Jin,Q., Shen,Y., Wang,J.H., Liu,H., Yang,J., Yang,F., Zhang,X.B., Zhang,J.Y., Yang,G.W., Wu,H.T., Dong,J., Sun,L.L., Xue,Y., Zhao,A.L., Gao,Y.S., Zhu,J.P., Chen,S.X., Yao,Z.J., Wang,Y., Lu,W.C., Qiang,B.Q., Wen,Y.M. and Hou,Y.D.  
TITLE Direct Submission  
JOURNAL Submitted (08-MAY-2001) Microbial Genome Center of Chinese Ministry of Public Health, 100 Yingxin Jie, Xuanwu Qu, Beijing 100052, P.R. China

FEATURES  
source location/Qualifiers  
1..12322  
/organism="Shigella flexneri 2a str. 301"  
/mol\_type="genomic DNA"  
/strain="301"  
/serotype="2a"  
/db\_xref="taxon:198214"  
218..1840  
gene

/gene="nadB"  
/locus\_tag="SF2636"  
/pseudo  
218..1840  
/gene="nadB"  
/locus\_tag="SF2636"  
/note="similar to Escherichia coli K12 quinolinate synthetase, B protein gi: 1788928 (541 aa). BLAST with identity of 97% in 541 aa. This CDS contains an in-frame stop codon. The sequence has been checked and is believed to be correct."

/pseudo  
/codon\_start=1  
/transl\_table=1  
complement(1825..2682)  
/gene="yfiC"  
/locus\_tag="SF2637"  
complement(1825..2682)  
/gene="yfiC"  
/locus\_tag="SF2637"  
/note="Residues 1 to 285 of 285 are 98 pct identical to residues 1 to 285 of a 285 aa protein from Escherichia coli K12 ref: NP\_417070.1"

/codon\_start=1  
/transl\_table=1  
/product="putative enzyme"  
/protein\_id="AAN44133.1"  
/db\_xref="GI:24053019"

/translation="MMGADYRRYGCNHLFWWISLGYCSHPVLSPFLOGFIFMSQSTSVLRNFTFKQFVAHDCAMKVGTDGILLGAWAPVAGVRCLDIGAGSLALMLAQRISDSVIIDAVELESEAAQAQENINQSPWAERINVTADIQOWITQGTFRDLITSNPYYOQGVECATPQREQARYTTLLDPSLITCAECITEEGFCVVLPEQINGFT ELALSMGWHLRLRTDVAENEARLPHRVLLAFSPQAGECFSDRLVIRGPDQNYSEAYTA LTOAFYLFM"  
2694..4028  
/gene="srnB"  
/locus\_tag="SF2638"  
2694..4028  
/gene="srnB"  
/locus\_tag="SF2638"  
/note="Residues 1 to 444 of 444 are 100 pct identical to residues 1 to 444 of a 444 aa protein from Escherichia coli O157:H7 ref: NP\_311469.1"

/codon\_start=1  
/transl\_table=1  
/product="ATP-dependent RNA helicase"  
/protein\_id="AAN44134.1"  
/db\_xref="GI:24053020"

/translation="MTVTTFSELELDESLEALQDKGFTRPTAIOAAIIPALDGRV LGSAPGTGKTAAYLLPALQHLDFEPKKSPPRILITPTRELAMQVADHARELAKH THLDIATITGGVAAYMNAEVSFNODIVATGRLQYKEENFDCRAVETLLDEAD RMLDMGFAQDIEHAGETRMKQTLLEFSATLEGAIDQAEERLLEDPVEVSANPSTRE RKKHQWYRRADBLEHKTALLVHLKQPEATRSIVFKRERVHELANWLREAGINNC YLEGEWYQGRNEAIKRLTEGRVNVIVATDVAARGIDIPVSHVENFMDPRSGDTYLH RIGRTARAQRKGTALISLVEADHLLLGKVGRIIEEPKARVIDELRPKTRAPSEKQIG KPSKYLAKRAEKKKAKEKEKPRYKKRHDTKNIGKRRKSGTGVPQTTEE"  
complement(4063..4989)  
/gene="yfiE"  
/locus\_tag="SF2639"  
complement(4063..4989)  
/gene="yfiE"

/locus\_tag="SF2639"  
/note="Residues 1 to 306 of 308 are 96 pct identical to residues 1 to 306 of a 308 aa protein from Escherichia coli O157:H7 EDL933 ref: NP\_289135.1"  
/codon\_start=1  
/transl\_table=1  
/product="putative transcriptional regulator LYSR-type"  
/protein\_id="AAN44135.1"

/db\_xref="GI:24053021"  
/translation="MFLSKPDIIDFRFGVMDLRRFTILKTVVEEGSFRLASQKLCTQSTVFPHIQOLEQSFVQLFEKIGRMKCLTREGKLLPHIYELTRVMDTLREAAKESD PDGELRVVSGETLLSYRMPQVLOFRFRAPKVRLSQLNCTYVIRDALINDEADVGVF



YRVGNDALNRRELGEQSLVLVASPOLADVDFTPEGRHNACSFINEPOCVEROIFES  
TLRORITVENTIELLISIESIKRCVANIGSVLYLPHFAVEKEIESGELIELPGEOSQ  
TITAMCAHYAGKAVSPAMHIFIQVCECFVA"

gene  
5047. .5634  
/gene="yfiK"  
/locus\_tag="SF2640"  
5047. .5634  
/gene="yfiK"  
/locus\_tag="SF2640"  
/note="Residues 1 to 195 of 195 are 98 pct identical to  
residues 1 to 195 of a 195 aa protein from Escherichia  
coli K12 ref: NP\_417073.1"  
/codon\_start=1  
/transl\_table=11  
/product="orf, conserved hypothetical protein"  
/protein\_id="AAN44136.1"  
/db\_xref="GI:24053022"  
/translation="MTPILLSAFWYTYLTITAMTPGPNNILALSSATSHGFRQSTRVLA  
GMSLGLIVMLCAGISFSLAVIDPAAVHLWSAGAIVIVLAWKIASPTKEDGLQA  
KPISEWASPALQFVNKIIILYGTALSTFVLPTQALSWVGVSVLLAMIGTFGNVCW  
ALAGHLFORLFRQYGRQNLNIVLALLVCAVRIFY"  
complement(5690. .6091)  
/gene="yfiD"  
/locus\_tag="SF2641"  
complement(5690. .6091)  
/gene="yfiD"  
/locus\_tag="SF2641"  
/note="Residues 7 to 133 of 133 are 99 pct identical to  
residues 1 to 127 of a 127 aa protein from Escherichia  
coli O157:H7 ref: NP\_311472.1"  
/codon\_start=1  
/transl\_table=11  
/product="putative formate acetyltransferase"  
/protein\_id="AAN44137.1"  
/db\_xref="GI:24053023"  
/translation="MPGRHMITGCIQITKAANADLNSFWLIDSEKGEARCIYAKAGY  
AEDEVAVASKLDIEYREVPVEYEGQHNLNVLRRELEDAVKHPEKYPQL  
TIRVSGYAVRNSLTPEQQRDVARTFTESL"  
6378. .6929  
/gene="ung"  
/locus\_tag="SF2642"  
6378. .6929  
/gene="ung"  
/locus\_tag="SF2642"  
/note="Residues 1 to 183 of 183 are 100 pct identical to  
residues 1 to 183 of a 229 aa protein pdb: 1EUG"  
/codon\_start=1  
/transl\_table=11  
/product="Chain A, Crystall Structure Of Escherichia Coli  
Uracil Dna Glycosylase And Its Complexes With Uracil And  
Glycerol: Structure And Glycosylase Mechanism Revisited"  
/protein\_id="AAN44138.1"  
/db\_xref="GI:24053024"  
/translation="MANELTWHDVLAEBKQPYFLNTLOTVASEROSGVTIYPQKDV  
FNAFRFTELDVKVVIIGODPYHGPQAHGLAFSVRPGIAIPSLNMYKELENTIPG  
FTRPNHGYLESWARQGVLLNTLTVRAQASHASLGMETFTDKVISLINQHREGVV  
FLWGSNAQKKGALIDKQRHVL"  
complement(7115. .8227)  
/gene="yfiF"  
/locus\_tag="SF2643"  
complement(7115. .8227)  
/gene="yfiF"  
/locus\_tag="SF2643"  
/note="Residues 1 to 355 of 370 are 97 pct identical to  
residues 4 to 358 of a 445 aa protein from Escherichia  
coli dbj: BAA16467.1"  
/codon\_start=1  
/transl\_table=11  
/product="orf, conserved hypothetical protein"  
/protein\_id="AAN44139.1"  
/db\_xref="GI:24053025"  
/translation="MLRNHGEPRGLRVICALRALVEKSNDEMKGKSGKVKVMYRS  
DDSDKRTHNPRGTGGGRPAKSRADGRRPARDDKOSQPRDRKEDSPMRTVSRAPG

DETEPKADHSGISGKSFIDPEVLARQRAEETRVYGENACQALFQSRPEALVRWFIQS  
VTPRFKALRWMAANRAKHVHVEBELTKASGETEHGVCFLIKRNGTTVOQWVIOA  
GAQDCYLALENESNPHNLGWMRSKAHFGVKGVVQDAALESQAIRTAEGAEHVQ  
PITGHNIWNLHDFRQAGYSVWTSSEQGPLFKTSLPAKRVLVLGQEVGELPDAARD  
PNDLRVKIDGTGNVAGLNISVATGVLLGEMWRONKA"

gene  
8359. .8778  
/gene="trxC"  
/locus\_tag="SF2644"  
8359. .8778  
/gene="trxC"  
/locus\_tag="SF2644"  
/note="Residues 1 to 139 of 139 are 100 pct identical to  
residues 1 to 139 of a 139 aa protein from Escherichia  
coli O157:H7 EDL933 ref: NP\_289141.1"  
/codon\_start=1  
/transl\_table=11

CDS

Query Match 99.4%; Score 745.2; DB 1; Length 12322;  
Best Local Similarity 99.6%; Pred. No. 1e-222;  
Matches 747; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GATCCATAACCCCAAACCTATCGAAATATCGAATCTAGAATATATAAATTCATTTT 60  
DB 4938 GATCCATAACCCCAAACCTATCGAAATATCGAATTTAGAATATAAACAATTCATTTT 4997

QY 61 TTAATGTTCCGTGTCGGGTACTGTCTACCAAAAACAGAGAGATACAGTGAACCCGAC 120  
DB 4998 TTAATGTTCCGTGTCGGGTACTGTCTACCAAAAACAGAGAGATACAGTGAACCCGAT 5057

QY 121 CCTTTAAGTCTTTTGGACTTACACCCCTGAATACCGCTATGACGCCACGAACCA 180  
DB 5058 CCTTTAAGTCTTTTGGACTTACACCCCTGAATACCGCTATGACGCCACGAACCA 5117

QY 181 TATTCGCCCTTAGCTCTGCTACGTGCGATGGAATTCGTCAAAAGTACCCGCTGCGC 240  
DB 5118 TATTCGCCCTTAGCTCTGCTACGTGCGATGGAATTCGTCAAAAGTACCCGCTGCGC 5177

QY 241 AGGATGAGTGTGGGATTTTGTATGTGATGTTACTGTGTGCGGCAATTCATTTTCACT 300  
DB 5178 AGGATGAGTGTGGGATTTTGTATGTGATGTTACTGTGTGCGGCAATTCATTTTCACT 5237

QY 301 GGCAGTATGACCCGCGACGCGTACCTTTTGAGTTGGCGGCGGCGCATATATTGT 360  
DB 5238 GGCAGTATGACCCGCGACGCGTACCTTTTGAGTTGGCGGCGGCGCATATATTGT 5297

QY 361 CTGCGTGGCGGAAATCGCCACCAAGCCCAACAAGAGACGGAATTCAGGCAAAACC 420  
DB 5298 CTGCGTGGCGGAAATCGCCACCAAGCCCAACAAGAGACGGAATTCAGGCAAAACC 5357

QY 421 AATCAGCTTTGGGCCAGCTTGTGTCAGATTTGTGAACGTCAAAATCATTTTGTACGG 480  
DB 5358 AATCAGCTTTGGGCCAGCTTGTGTCAGATTTGTGAACGTCAAAATCATTTTGTACGG 5417

QY 481 TGTACGGCACTGCGACGTTGTGTCGCGCAAAACACAGGCGTTAAGCTGGGTAGTTGG 540  
DB 5418 TGTACGGCACTGCGACGTTGTGTCGCGCAAAACACAGGCGTTAAGCTGGGTAGTTGG 5477

QY 541 CGTCAGCGTTTGTGCGGATGATGCGGACGTTTGGCAATGTGTGCGGCGCTGCGCGG 600  
DB 5478 CGTCAGCGTTTGTGCGGATGATGCGGACGTTTGGCAATGTGTGCGGCGCTGCGCGG 5537

QY 601 GCATCTGTTCAAGCATGTTTCCGCAATATGTCGCCAGTTAATATCGTGTGCCCT 660  
DB 5538 GCATCTGTTCAAGCATGTTTCCGCAATATGTCGCCAGTTAATATCGTGTGCCCT 5597

QY 661 GTTGTGCTATATGCGCGGTACGCAATTTCTATTAACGAAAAAAGCGGAAGAGTCCG 720  
DB 5598 GTTGTGCTATATGCGCGGTACGCAATTTCTATTAACGAAAAAAGCGGAAGAGTCCG 5657

QY 721 CCTCTTCGCTTAAGTACTTGTCTACTTAAG 750  
DB 5658 CCTCTTCGCTTAAGTACTTGTCTACTTAAG 5687

RESULT 7	
LOCUS AE016987	290380 bp DNA linear BCT 22-APR-2003
DEFINITION Shigella flexneri 2a str. 2457T section 10 of 16 of the complete genome.	
ACCESSION AE016987 AE014073	
VERSION AE016987.1 GI:30042162	
KEYWORDS	
SOURCE ORGANISM	Shigella flexneri 2a str. 2457T Shigella flexneri 2a str. 2457T Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Shigella. . (bases 1 to 290380)
REFERENCE AUTHORS	Wei,J., Goldberg,M.B., Burland,V., Venkatesan,M.M., Deng,W., Fournier,G., Mayhew,G.F., Plunkett,G. III, Rose,D.J., Darling,A., Mau,B., Perna,N.T., Payne,S.M., Runyen-Janecky,L.J., Zhou,S., Schwartz,D.C. and Blattner,F.R. Complete Genome Sequence and Comparative Genomics of Shigella flexneri Serotype 2a Strain 2457T Infect. Immun. 71 (5) , 2775-2786 (2003)
TITLE JOURNAL PUBMED	2 (bases 1 to 290380) Wei,J., Goldberg,M.B., Burland,V., Venkatesan,M.M., Deng,W., Fournier,G., Mayhew,G.F., Plunkett,G. III, Rose,D.J., Darling,A., Mau,B., Perna,N.T., Payne,S.M., Runyen-Janecky,L.J., Zhou,S., Schwartz,D.C. and Blattner,F.R. Direct Submission Submitted (13-JUN-2002) Genetics Laboratory, University of Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES	Location/Qualifiers 1..290380 /organism="Shigella flexneri 2a str. 2457T" /mol_type="genomic DNA" /strain="2457T" /serotype="2a" /db_xref="taxon:198215" complement(82..1209) /gene="gcpE" /locus_tag="S2733" complement(82..1209) /gene="gcpE" /locus_tag="S2733" notes="residues 1 to 372 of 375 are 96.23 pct identical to residues 1 to 372 of 372 from Escherichia coli K-12 : B2515"
gene CDS	/codon_start=1 /transl_table=1 /product="hypothetical protein" /protein_id="AAP17888.1" /db_xref="GI:30042163" /training="MANOAPIQRKSTRIVGVNPIGDGAPIAVQSMTRRTDVEATVNOIKALERVADIVRSVPTMDAAEFKLKORVNPLVADIHFDYRIALKVAEGVWDLRLNPNGINGNEERIRMVVDCCARDKNIPIRIGVNAAGSLKDLOEKYGEPYQALLSEAMRHVDHLDRLNFDQFKSVSKASDVFLAVESYRLAKOIDPLHGITEAGARSGAVKSAIGLGLISEGIDTLRVSLAADPVEBIKYGFILKSLRIRSRGINFIACPCSRQC EFDVIGTVNALEGRLEDLIITPMDSIIGCVANGGEALVSTLVGTGNKKSGHYEDGV RKDRLDNNDMIDQLEARIRAKASQLDEARRIDVQVGKIIT" complement(1236..2249) /gene="yfgA" /locus_tag="S2734" complement(1236..2249) /gene="yfgA" /locus_tag="S2734" /function="putative membrane; Not classified" /notes="residues 1 to 337 of 337 are 80.71 pct identical to residues 1 to 337 of 337 from Escherichia coli K-12 : B2516"
CDS	/codon_start=1 /transl_table=1 /product="putative membrane protein" /protein_id="AAP17889.1" /db_xref="GI:30042164"

```

gene
complement (2534 . .3688)
/gene="yfg8"
/locus_tag="S2735"
complement (2534 . .3688)
/gene="yfg8"
/locus_tag="S2735"
/notes="residues 1 to 384 of 384 are 99.47 pct identical to
residues 1 to 384 of 384 from Escherichia coli K-12 :
B2517"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAP17890.1"
/db_xref="GI:30042165"
/translacion="MSEQLVTPENVTTKDGKINLLDLNRQMRFPKDGKTERADQ
VKMVMHYCCDNFDEMTDINKVLRGKLEVAEIRAPEVEEQRSSDGTIKMAIVADQ
RVETVYIPEDDRATLCVSSQVGCALCEKFCSTAOQGFNRNLRVSEIIGYWRAKIVG
AAKVTGORPITNVVMGGEPLINNNVPMAMEIMLDFEGLSKRVTLSGVPA
LDKLDGMIDVALAISLHAPNDIIRDEIVIPINKYNIETFLAAYRYLEKSNANQGRVT
IEYMLDHNVDGTEHAHQLAELIKDTPCKINLIPNNPEPDAPYGRSSNSRTRDRFSKVL
MSYGFTTIVRKTRGDDIDACGQLAGVIDRTRKRTLRRKMQGEAIDIKAV"
complement (3838 . .4269)
/gene="ndk"
/locus_tag="S2736"
complement (3838 . .4269)
/gene="ndk"
/locus_tag="S2736"
/function="enzyme; Purine ribonucleotide biosynthesis"
/notes="residues 1 to 143 of 143 are 100.00 pct identical
to residues 1 to 143 of 143 from Escherichia coli K-12 :
B2518"
/codon_start=1
/transl_table=11
/product="nucleoside diphosphate kinase"
/protein_id="AAP17891.1"
/db_xref="GI:30042166"
/translacion="MAIERTPSITKPNAAKVNIGNIPARFEAGFKIVGTRMLHLTV
EQARGFYAEHDGKPPFDGLVEFMTSGPIVSVLEGENAVQHRDLLGATNPANALAGT
LRADYADSLTENGTHGSDSVESAAREIAYFFGEGEYCPRT"
complement (4418 . .6517)
/gene="pbcC"
/locus_tag="S2737"
/notes="truncated"
/pseudo
6514 . .7961
/notes="insertion element"
/insertion_seq="ISSf14"
complement (6534 . .7730)
/locus_tag="S2738"
complement (6534 . .7730)
/locus_tag="S2738"
/function="IS, phage, Tn; Transposon-related functions"
/notes="residues 1 to 398 of 398 are 90.95 pct identical to
residues 1 to 398 of 398 from GenPept :
>gb|AAK18438.1|AF348706.127 (AF348706) putative
transposase for IS110 [Shigella flexneri]"
/codon_start=1
/transl_table=11
/product="ISSf14 orf"
/protein_id="AAP17892.1"
/db_xref="GI:30042167"
/translacion="MTESSDYBSVQVFIGVDVGKDTTHHAVAINRSCKRLFDKALPNDB
NKLRSLSDLKOHGOILVVDOPATIGALPVAVARSEGVIVGLPGLAMRRIADLHAG
EAKTDVDRDAIITAEARTLPHALRTKLADEQIAELSMLCGFPDDDLAAQTQASNRHIF
GLTQIHPAERVLGPRLEHPAVLDLLQRYSPSEKLASLGEKLAALCKLAPRLGKX
LAADIAQALAEQTVVVPGETNAAVVLPRLALQLITLRQRDEVALEVEQVLAHPLYV

```

VLTSMPGVVRTAARLLTEVACRAFASAAHLAAVAGLAPVTRRSSGSSIRGEHPSRRGN  
KALKRALFLSAFAALRDLPSRAYYTRKMSQGRHNQALIALARRRCDVLFAMMRDGT  
YTPQGS"  
gene 7960. .8442  
/gene="ssea"  
/locus\_tag="S2740"  
/note="truncated"  
/pseudo  
complement(9261. .10037)  
gene /gene="sseb"  
/locus\_tag="S2741"  
complement(9261. .10037)  
CDS /gene="sbeb"  
/locus\_tag="S2741"  
/function="phenotype; Not classified"  
/note="residues 1 to 258 of 258 are 94.96 pct identical to  
residues 4 to 261 of 261 from Escherichia coli K-12 :  
B2522"  
/codon\_start=1  
/transl\_table=11  
/product="enhancer of serine sensitivity"  
/protein\_id="AAP17893.1"  
/db\_xref="GI:30042168"  
/translation="MSETKNELEDLLEKATEPAHRPAFRTLLESTVWVGTAQGE  
AVEDSALDLQHEKEDGTSVIPFTSLALQAVEDEQA FVMPVRTLFEMTLGETL  
FLNAKLPTGKEFMREISLILGEGNPLSSOEILGEGSLILSEVAEPPAQMIDSLTT  
LFKTIKPVKRAFICSIKENEAQPNLIGIADGDIEEIIQATGSVATDTLPGDEPID  
ICQVKGEKGI SHFTIEHIAFYERRMGFLRDFKQNRli"  
repeat\_region 10137. .10170  
/note="REP (repetitive extragenic palindromic) element,  
contains 1 REP sequence"  
complement(10179. .11462)  
gene /gene="pepB"  
/locus\_tag="S2742"  
complement(10179. .11462)  
CDS /gene="pepB"  
/locus\_tag="S2742"  
/function="putative enzyme; Not classified"  
/note="residues 1 to 427 of 427 are 92.50 pct identical to  
residues 30 to 456 of 456 from Escherichia coli K-12 :  
B2523"  
/codon\_start=1  
/transl\_table=11  
/product="putative peptidase"  
/protein\_id="AAP17894.1"  
/db\_xref="GI:30042169"  
/translation="MTEAMKITLSTOPADARWGEKATYSINNDGITLHNGADDLGLI  
QRARKIDGLIGIKVQLSGEGWDADRCWAFWQYKGPKTRKVEWPDLDAAQROELDN  
RLMIIDWRDITINAPAEELGPSOLAQRAVDLISNVAQRAVDLITKGEPLREQYMG  
HTVGRGSESPVLLALDYNPTGDEKAPVACTLVGKIGITPDSGGYSIKQTFMDSKMSD  
MGAATVTGALAPAIMRGLNKRVKLFLCCADNLISGNAFKLDIITVRNGKVEVMT  
DAEGRVLADGLIDASAQKPEMIIIDAATLTGAAKTALGNDYHALFSFDALAGRLLAS  
ASQENEPFWRPLPLAEFHRSQLPSNFAELNNTGSAAYPAGASTAGFLSHVENYQGGW  
LHIDCSATYRKAPVEQWSAGATGLGVRTIANDLTA"  
gene complement(11521. .11721)  
/gene="yfhJ"  
/locus\_tag="S2743"

Query Match 99.4%; Score 745.2; DB 1; Length 290380;  
Best Local Similarity 99.6%; Pred. No. 2.1e-222;  
Matches 747; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GATCCATAACCCCAAACTATCGAAATATCGAATCTAGAAATATAAAACATTCAATTTT 60  
|||||  
Db 82162 GATCCATAACCCCAAACTATCGAAATATCGAATTTAGAAATATAAAACATTCAATTTT 82221  
QY 61 TTAATGTTCCGTGCGGTACTGTCTACCAAAACAGAGAGATTAACAAGTGACACCGAC 120  
|||||  
Db 82222 TTAATGTTCCGTGCGGTACTGTCTACCAAAACAGAGAGATTAACAAGTGACACCGAT 82281  
QY 121 CCTTTAAGTGCTTTTGGACTTACACCCCTGATTAACCGCTATGACGCCAGACCGAACAA 180  
|||||  
Db 82282 CCTTTAAGTGCTTTTGGACTTACACCCCTGATTAACCGCTATGACGCCAGACCGAACAA 82341

QY 181 TATTCGCCCCCTTAGCTCTGCTACTACGTGCATGATTTTCGTCAAAGTACC CGCGCTGCGC 240  
|||||  
Db 82342 TATTCGCCCCCTTAGCTCTGCTACTACGTGCATGATTTTCGTCAAAGTACC CGCGCTGCGC 82401  
QY 241 AGGATGAGTCTGGGATTTTGTATGTGATGTTACTGCTGCGGCATTTCATTTTCACT 300  
|||||  
Db 82402 AGGATGAGTCTGGGATTTTGTATGTGATGTTACTGCTGCGGCATTTCATTTTCACT 82461  
QY 301 GGCAGTGAATGACCCCGCAGCGGTACACCTTTTGATGTGGCGGGCGGCATATAATTGT 360  
|||||  
Db 82462 GGCAGTGAATGACCCCGCAGCGGTACACCTTTTGATGTGGCGGGCGGCATATAATTGT 82521  
QY 361 CTGGCTGGCGTGAAAATCGCCACACCGCCCAACAAAGAAAGACGACTTCAGGCAAAACC 420  
|||||  
Db 82522 CTGGCTGGCGTGAAAATCGCCACACCGCCCAACAAAGAAAGACGACTTCAGGCAAAACC 82581  
QY 421 AATCAGCTTTGGGCCAGCTTTGCTTGTGAGTTTGCAACGTCAAATCATTTTGTACGG 480  
|||||  
Db 82582 AATCAGCTTTGGGCCAGCTTTGCTTGTGAGTTTGCAACGTCAAATCATTTTGTACGG 82641  
QY 481 TGTACGGCACTGTTCGACGTTTGTCTGCGGCAACACAGCGCTTAAGCTGGGTAGTTGG 540  
|||||  
Db 82642 TGTACGGCACTGTTCGACGTTTGTCTGCGGCAACACAGCGCTTAAGCTGGGTAGTTGG 82701  
QY 541 CGTCAGCGTTTGTGCGGATGATTTGGACGTTTGGCAATGTGTGCTGGCGCTGCGCGG 600  
|||||  
Db 82702 CGTCAGCGTTTGTGCGGATGATTTGGACGTTTGGCAATGTGTGCTGGCGCTGCGCGG 82761  
QY 601 GCATCTGTTACGCGATGTTTTCGCCACAGTATGTCGCCAGTTAAATATCGTGTGCCCT 660  
|||||  
Db 82762 GCATCTGTTACGCGATGTTTTCGCCACAGTATGTCGCCAGTTAAATATCGTGTGCCCT 82821  
QY 661 GTTGCTGCTATATTGCCCGCGTACGCATTTCATATTAACGAAAAAGCGGAAGGTCGC 720  
|||||  
Db 82822 GTTGCTGCTATATTGCCCGCGTACGCATTTCATATTAACGAAAAAGCGGAAGGTCGC 82881  
QY 721 CCTCTTCGCTTAGTAAGTGTCTACTTAAG 750  
|||||  
Db 82882 CCTCTTCGCTTAGTAAGTGTCTACTTAAG 82911

RESULT 8  
AE016764 300099 bp DNA linear BCT 24-DEC-2002  
LOCUS  
DEFINITION Escherichia coli CFT073 section 10 of 18 of the complete genome.  
ACCESSION AE016764 AE014075  
VERSION AE016764.1 GI:26109116  
KEYWORDS  
SOURCE  
ORGANISM Escherichia coli CFT073  
REFERENCE  
AUTHORS Escherichia coli CFT073  
1 (bases 1 to 300099)  
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
Enterobacteriaceae; Escherichia.  
TITLE  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 17020-17024 (2002)  
PUBMED 12471157  
REFERENCE  
AUTHORS 2 (bases 1 to 300099)  
Welch,R.A., Burland,V., Plunkett,G.D. III, Redford,P., Roesch,P.,  
Rasko,D.A., Buckles,E.L., Liou,S.-R., Boutin,A., Hackett,J.,  
Stroud,D., Mayhew,G.F., Rose,D.J., Zhou,S., Schwartz,D.C.,  
Perna,N.T., Mobley,H.L.T., Domnenberg,M.S. and Blattner,F.R.  
Extensive Mosaic Escherichia coli  
of Uropathogenic Escherichia coli  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 17020-17024 (2002)  
TITLE  
JOURNAL Direct Submission  
Submitted (20-JUN-2002) Genetics Laboratory, University of  
Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA  
FEATURES  
source  
1. .300099  
/organism="Escherichia coli CFT073"



gene /mol\_type="genomic DNA"  
/strain="CFT073"  
/db\_xref="taxon:199310"  
complement(163. .1383)  
/gene="fabB"  
/locus\_tag="c2869"  
complement(163. .1383)  
/gene="fabB"  
/locus\_tag="c2869"  
/EC\_number="2.3.1.41"  
/function="enzyme; Fatty acid and phosphatidic acid biosynthesis"  
/note="Escherichia coli K-12 ortholog: b2323; Escherichia coli O157:H7 ortholog: z3586"  
/codon\_start=1  
/transl\_table=1  
/product="3-oxoacyl-[acyl-carrier-protein] synthase I"  
/protein\_id="AAN81319.1"  
/db\_xref="GI:26109117"  
/translation="MKRAVITGLGVSSIGNNQEVLASLREGSGITFSQELKDSGM RSHVWGNVKLDTGLIDRKVRFMSDASIYAFLSMEQAIADAGLSPAYQNNPRVGLI AGSGGSGPRQVFGADAMRGPRLKAVSPYVVTKAMASGVSACLAPEFKIHGVNYSIS SACATSAHCIGNAVEQIOLGKODIVFAGGSEELCWEMACEFDAMGALSTKYNDPEKA SRTYDAHRDGFVIAGGGGVVVEELEHALARGAHIYAEIIVGYGATSDGADMVAPSGEG AVRCKMAMHGVDPIDILNSGTSTPYGDVKEIAIREFVGDKSPAISATKAMTGHSLGAAGVQEAITYSLMLEHGFIAPSINIEELDEQAAGLNIYETTDRELTVMSNSFGF GGTNATLVMRKLD"  
1482. .3548  
/gene="yfcK"  
/locus\_tag="c2870"  
1482. .3548  
/gene="yfcK"  
/locus\_tag="c2870"  
/function="putative"  
/note="Escherichia coli K-12 ortholog: b2324; Escherichia coli O157:H7 ortholog: z3587"  
/codon\_start=1  
/transl\_table=1  
/product="Hypothetical protein yfcK"  
/protein\_id="AAN81320.1"  
/db\_xref="GI:26109118"  
/translation="MRSLTTHAKIEPPVRRVTVVKHYSIQPANLENNAEGTPVSRDF DDVYFSNDNGLEETRYVFLGKNHLEARPEHPPLFVVAESGFGTGLNFLTLMQAFDQ FREAHPOAQLORLHFISFEKPLTRADILAHQHPLELAPMAEQLOQWPLPLPGCHR LLLDEGHVTLDLMEGDLNELTSQDLSLNOKYDAWELDGFAPAKNPDWTONLENAMA RLARPGSTLATFTSAGFVRKGLQEAGFTWQKRGFKRKREMLCGVMEOTLPPFCSTPW FNRMGSNKQEAALIGGGIASALLSLALLRGMQVTLYCADEAPALGASGNRQALYPL LSKHDEALNRFNSNAFTFARFYDLVPKFDHWCQVTOLGWDEKSOHKIAQMLSMDL PAELAVAVEANAVEQITGVATNCSGITYPQGGWLCPAELTRNVLELAQQOGLQIHYYQ OLODLSRKDCGWLNFAGDOHATSHVVLANGHOISGFSOTSLPYSVAGQVSHIPT TPELAKLKQVLCYDGYLTPQNAPANOHCIASYHRGSEETAYSDEDOQNRQRLIDCF PQVQWAKEVDVSGKEARCGVCATRDHLLPMGVNVPDYDATLVEYASLAEKKDEAVSAP VYDDLFMFAALGSRGLCSAPLCAEILAAQWSEEPFPMDASTLAALNPNRLWVRKLLKG KAVKAG"  
complement(3669. .3947)  
/gene="yfcL"  
/locus\_tag="c2871"  
complement(3669. .3947)  
/gene="yfcL"  
/locus\_tag="c2871"  
/note="Escherichia coli K-12 ortholog: b2325; Escherichia coli O157:H7 ortholog: z3588"  
/codon\_start=1  
/transl\_table=1  
/product="Hypothetical protein yfcL"  
/protein\_id="AAN81321.1"  
/db\_xref="GI:26109119"  
/translation="WIAEFESRIALLIDGMVDHASDDELFASGYLGRHLLTLLIAELLES GDDHSAQAVHTTVSQSLKAIAGAGELSPRDQALVTDMWEHLFQOASQO"  
complement(3981. .4529)  
/gene="yfcM"  
/locus\_tag="c2872"  
complement(3981. .4529)

gene /gene="yfcM"  
/locus\_tag="c2872"  
/function="putative"  
/note="Escherichia coli K-12 ortholog: b2326; Escherichia coli O157:H7 ortholog: z3589"  
/codon\_start=1  
/transl\_table=1  
/product="Hypothetical protein yfcM"  
/protein\_id="AAN81322.1"  
/db\_xref="GI:26109120"  
/translation="MNSTHYEQLEIFNSCFADENFTRLIKGDDEPIYLPADAEPVY NRIYFAHGFYASAIHEISHWCIAKARLDLVDFGYWYCPDGRDAQOSQFEDVEVKPQ ALDMLFCVAAGYPFNVSCDNLEGDPEPDRVVFORRVHAQWMDYLANGIPERPARFIKA LQNYHTPELTAEQFPWPEALN"  
complement(4529. .5338)  
/gene="yfcA"  
/locus\_tag="c2873"  
complement(4529. .5338)  
/gene="yfcA"  
/locus\_tag="c2873"  
/translation="putative structure"  
/note="Escherichia coli K-12 ortholog: b2327; Escherichia coli O157:H7 ortholog: z3590"  
/codon\_start=1  
/transl\_table=1  
/product="Hypothetical protein yfcA"  
/protein\_id="AAN81323.1"  
/db\_xref="GI:26109121"  
/translation="METFNSLFWVSPLLGLVLFVAMLAGFIDSIAGGGLTTPALM AAGMSPANALATNKLQACGSGISATIFYIRRKVVSISDQKLNIAMTFVGSMSGALLVQ YVQADVLRQILPILVICIGLVFLMLPKEEDRORRMVGLPEALIAGGCVGIFYDGFPG PAAGSFYALAFVTLGCFNLAKATAHAKLNLATSNIGLLLEILGKVIWATGFVMLVG QFLGARMGSRVLVLSKGQKLIRPMIVIVSAVMSAKLLYDSHQEILHMLGMN"  
complement(5338. .6162)  
/gene="mepA"  
/locus\_tag="c2874"  
complement(5338. .6162)  
/gene="mepA"  
/locus\_tag="c2874"  
/EC\_number="3.4.99.-"  
/function="enzyme; Murein sacculus, peptidoglycan"  
/note="Escherichia coli K-12 ortholog: b2328; Escherichia coli O157:H7 ortholog: z3591"  
/codon\_start=1  
/transl\_table=1  
/product="Penicillin-insensitive murein endopeptidase precursor"  
/protein\_id="AAN81324.1"  
/db\_xref="GI:26109122"  
/translation="MNKTATALLALLASSASLAATPWOKITOPVPGSAOSIGSPSNGC IVGADTLPQSEHYQVMRTDQRRYFGHPDLVMFIQRLSRQVSNLGMGTVLIGDMGMPA GGRFNGGASHQTLGDVDIFLOLPKTRWTSQOLLRPOALDIVSRDGKHWVPALMKPEI FSLIKLAADKDVTRI FVNPAIKQQLCLDAGTDRDWLRKVRPWFQRAHMHVRLRCPA DSIKECEDQPLPPGDGCAELQSWFEPKPGTTKPEKTPPPLPSPSCALLDBHVI"  
complement(6166. .7254)  
/gene="aroc"  
/locus\_tag="c2875"  
complement(6166. .7254)  
/gene="aroc"  
/locus\_tag="c2875"  
/EC\_number="4.2.3.5"  
/function="enzyme; Amino acid biosynthesis: Chorismate"  
/note="Escherichia coli K-12 ortholog: b2329; Escherichia coli O157:H7 ortholog: z3592"  
/codon\_start=1  
/transl\_table=1  
/product="Chorismate synthase"  
/protein\_id="AAN81325.1"  
/db\_xref="GI:26109123"  
/translation="MMAGNTIGQLFRVTTFGESHGLALGCIYDGVBPGLPTEADLOH DLDRRPGTSRYTTQREPPDQKILSGVEFEGVTTGTSIGLLIENTDQRSQDYSAIKDV FRPGHADITYEQKYGRLRGGGRSSARETAMRVAAAGALAKKYLAEEKGIEIRGCLTQ MGDIPLEIKDWSQVEQNPFPCPDPIXDALDELMRALKKEGSIKAKVTVAASGVPAQ



LGEVPFDRLDADIAHAlMSINAVKVEIGDGEDVVALRGSQNRDEITKDFQSNHAG  
ILGSISSGOOIIAHMALKPTSSITVPERITNRFGEVEEMITKGRHDPVGIKRAVPIAE  
AMLAIVMDHLRORAQONADVKTDIRW"  
complement (7286 .8551)  
/gene="yfcB"  
/locus\_tag="c2876"  
complement (7286 .8551)  
/gene="yfcB"  
/locus\_tag="c2876"  
/EC\_number="2.1.1.72"  
/function="putative"  
/note="Escherichia coli K-12 ortholog: b2330; Escherichia  
coli O157:H7 ortholog: 23593"  
/codon\_start=1  
/transl\_table=11  
/product="Hypothetical adenine-specific methylase yfcB"  
/protein\_id="AAN81326.1"  
/db\_xref="GI:26109124"  
/translation="MAGIGLFLNOPYWHPADFTLRSVDNRVLNLASPRHOLAKOR  
LVLLAKCFRLFHLSFSISNPTVYPRSTIDRFKRIFFHSPRVNADLSSSSWQTSRRN  
YCDHALBEEYVDKIFVDEAVNLSQFDMLRMSVSRSANINRGTDGNDPDEAVOLV  
LPSLYLPDIPEDMRTARLTSSSEKRIVERVIRVRNERIPAYLTNRAMFEGHEPYVD  
ERVLVPRSPDIGELINNKFAGLSDMCTGSCGSCIALACAYAFPEAEVDADVIS  
PDALAIAEONIIEHGLIHNVIPRSDFRDLPKVQYDLITNPYPVDAEDMSDLPNEX  
RHEPELGASGTDGLKTRILIGNADYLDADDGVLICEVGSNMVHLMEQYPDVPTWL

Query Match 98.5%; Score 738.8; DB 1; Length 300099;  
Best Local Similarity 99.1%; Pred. No. 2.2e-220;  
Matches 743; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 GATCCATAACCCCAAACCTATCGAAAATATCGAATCTAGAATATAAAACATTCATTTT 60  
DB 255985 GATCCATAACCCCAAACCTATCGAAAATATCGAATCTAGAATATAAAACATTCATTTT 256044

QY 61 TTAATGTTCCGTGCGGGTACTGCTACTACCAAAACAGAGAGATAACAAGTACACCGAC 120  
DB 256045 TTAATGTTCCGTGCGGGTACTGCTACTACCAAAACAGAGAGATAACAAGTACACCGAC 256104

QY 121 CCTTTAAGTCTTTTGGACTTACACCCGTGATTACCGCTATGACGCCAGACCGAACAA 180  
DB 256105 CCTTTAAGTCTTTTGGACTTACACCCGTGATTACCGCTATGACGCCAGACCGAACAA 256164

QY 181 TATTCGCCCTTACTCTGCTACGCGCATGATTTGCTCAAAGTACCCCGTGTGGC 240  
DB 256165 TATTCGCCCTTACTCTGCTACGCGCATGATTTGCTCAAAGTACCCCGTGTGGC 256224

QY 241 AGGATGAGTCTGGGATTTTGTATGTGATGTTACTGTGTGCGGCATTTCATTTCACT 300  
DB 256225 AGGATGAGTCTGGGATTTTGTATGTGATGTTACTGTGTGCGGCATTTCATTTCACT 256284

QY 301 GGCAGTATGACCCGGCAGCGGTACACCTTTGAGTTGGCGGGCGGCATATATTGT 360  
DB 256285 GGCAGTATGACCCGGCAGCGGTACACCTTTGAGTTGGCGGGCGGCATATATTGT 256344

QY 361 CTGGCTGCGTGAATAATCGCCACGACCCCAACAAGAAGACGACTTCAGGCCAAACC 420  
DB 256345 CTGGCTGCGTGAATAATCGCCACGACCCCAACAAGAAGACGACTTCAGAACAAACC 256404

QY 421 AATCAGCTTTGGGCGAGCTTGTGCTTGCAGTTTGTGAACGTCAAATCATTTTGTACGG 480  
DB 256405 AATCAGCTTTGGGCGAGCTTGTGCTTGCAGTTTGTGAACGTCAAATCATTTTGTACGG 256464

QY 481 TGTACGGGCACTGTGACGTTGTTCTGCCGCAACACAGGCGTTAAGCTGGGGTAGTTGG 540  
DB 256465 TGTACGGGCACTGTGACGTTGTTCTGCCGCAACACAGGCGTTAAGCTGGGGTAGTTGG 256524

QY 541 CGTCAGCGTTTGTGCTGGCGATGATGGGACGTTTGGCAATGTGTGCTGGGCGCTGGCGGG 600  
DB 256525 CGTCAGCGTTTGTGCTGGCGATGATGGGACGTTTGGCAATGTGTGCTGGGCGCTGGCGGG 256584

QY 601 GCATCTGTTTACGCGATTGTTTGGCCAGTATGCTGCCAGTTAATATCGTCTGGCCCT 660  
DB 256585 GCATCTGTTTACGCGATTGTTTGGCCAGTATGCTGCCAGTTAATATCGTCTGGCCCT 256644

QY 661 GTTGTGCTATTATTCGCCGGGTACGCATTTTCTATTACGAAAAAAGCCGAGAGAGTCCG 720  
DB 256645 GTTGTGCTATTATTCGCCGGGTACGCATTTTCTATTACGAAAAAAGCCGAGAGAGTCCG 256704

QY 721 CCTCTCCGCTTAGTACTTGCTACTTAAG 750  
DB 256705 CCTCTCCGCTTAGTACTTGCTACTTAAG 256734

RESULT 9  
AE005488  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
FEATURES  
source

AE005488 12337 bp DNA linear BCT 21-MAR-2001  
Escherichia coli O157:H7 EDL933 genome, contig 3 of 3, section 107  
of 290.  
AE005488 AE005174  
AE005488.1 GI:12516994  
Escherichia coli O157:H7 EDL933  
Escherichia coli O157:H7 EDL933  
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
Enterobacteriaceae; Escherichia.  
1 (bases 1 to 12337)  
Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,  
Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,  
Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,  
Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,  
Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,  
Welch,R.A. and Blattner,F.R.  
Genome sequence of enterohaemorrhagic Escherichia coli O157:H7  
Nature 409 (6819), 529-533 (2001)  
21074935  
11206551  
2 (bases 1 to 12337)  
Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,  
Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,  
Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,  
Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,  
Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,  
Welch,R.A. and Blattner,F.R.  
Direct Submission  
Submitted (22-OCT-2000) Laboratory of Genetics, University of  
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA  
Location/Qualifiers  
1. 12337  
/organism="Escherichia coli O157:H7 EDL933"  
/mol\_type="genomic DNA"  
/strain="EDL933"  
/serotype="O157:H7"  
/db\_xref="taxon:155864"  
/note="enterohemorrhagic"  
235. 1857  
/gene="nadB"  
/note="synonym: 23856"  
235. 1857  
/gene="nadB"  
/function="enzyme; Biosynthesis of cofactors, carriers:  
Pyridine nucleotide"  
/note="Residues 1 to 540 of 540 are 99.07 pct identical to  
residues 1 to 540 of 540 from Escherichia coli K-12 Strain  
MG1655: B2574"  
/codon\_start=1  
/transl\_table=11  
/product="quinolinate synthetase, B protein"  
/protein\_id="AAG57690.1"  
/db\_xref="GI:12516995"

CDs  
gene  
CDS

STFYAOGGIAAVFDETDSDSHVEDTLIAGAGICDRAVEFVASNARSCVOMLIDGV  
LFDTHIQNGEESYHLTRGEGSHRRIHAADATGEVETLVSKALNHPNIRVRS  
NAVDLIISDKIGLPGTRVVGAVMNRNKEVETCHAKAVVLATGASKVQYTNPD  
ISSGDIAAMARAGCRVANLBENOHPALYHPOANPLLEALRGEGAYLKRPGTR  
EMPDFVRGELAPRDIVARAIDHEMKRLGADCMFLDISHKPADFIROHPMIYEKILG  
LGIDLTOEVPPIVPAHYTCGGVMVDDHGRTDVEGLYAIGEVSYTGLHGANRMA NSL

LECLVGSWSAEDITRRMPYADISTLPWDESRVENPDERVLIQNMWHELRLFMWDY  
VGIVRTTKRLERLRITMLOQEIIDEYYAHFRVSNLLELRLVQVALLIVRCAMMRK  
ESRGHFTLDYPELLTHSGPSISGNHYINR"  
/complement (1842. .2699)  
/gene="yfiC"  
/note="synonym: Z3857"  
complement (1842. .2699)  
/gene="yfiC"  
/function="putative enzyme; Not classified"  
/note="Residues 1 to 285 of 285 are 95.08 pct identical to  
residues 1 to 285 of 285 from Escherichia coli K-12 Strain  
MG1655: B2575"  
/codon\_start=1  
/transl\_table=11  
/product="putative enzyme"  
/protein\_id="AAG57691.1"  
/db\_xref="GI:12516996"  
/translation="MMGADYRRYGCNLHLFWMISLGYCSHPVLSPEFLOGFIFMSOS  
TSVLRNGFTFKQFFVAHDCAMKYGTGILLGAMAPVAGVKRCLDIGAGSGLLALML  
AORTSDSVIIDAVELESEATQAENVAAQS PWLERINVHTADIQWVTOQTARFDLI  
SNPPYEQVECATPQEQARYTSDHQTLTCAECITEEGFCVVLPEQIGNSFT  
ELALSMGWHRLRLRTDVAENEARLPHRYLLAFSPAQECFSDRLVIRGPDQNYSEAYTA  
LTOAFYLEFM"  
/transl\_table=11  
/codon\_start=1  
/product="ATP-dependent RNA helicase"  
/protein\_id="AAG57692.1"  
/db\_xref="GI:12516997"  
/translation="MTVTTFSELDELDESLLLEALQDKGFTRPATIQAAIPPALDGRDV  
LGSAPGTGKTAAYLLPALQHLDFPRKSGPRILILPTRELAMQVADHARELAKH  
THLDIATITGVAAVMNHAEFSENDIVVATTGRLLOYIKEENFCRAVETLIXDEAD  
RMLDMGFAQDIEHIAGETRWKQTLFSATLEGDAIQDFAERLLEDPEVSANPSTRE  
RKKIHOWYRADDLLEHKTALVHLKQPEATRSIVFVRKREVRHMLNMLEAGINNC  
YLEGEMVQGRNEAIKRLTEGRVNLVATDVAAAGIDIPDVSHVFNPDMPRSGDTYLH  
RIGRTARAGRKGTALISLVEADHLLGKGVYIEBPIKARVIDELRPKTRAPSEKQTG  
KPSKRVLLAKRAEKKKAKEKEKPRVKKRHRDTKNIGKRRKPSGTVPPQTTEE"  
complement (4078. .5004)  
/gene="yfiE"  
/note="synonym: Z3860"  
complement (4078. .5004)  
/gene="yfiE"  
/function="putative regulator; Not classified"  
/note="Residues 1 to 308 of 308 are 96.75 pct identical to  
residues 1 to 308 of 308 from Escherichia coli K-12 Strain  
MG1655: B2577"  
/codon\_start=1  
/transl\_table=11  
/product="putative transcriptional regulator LYSR-t-type"  
/protein\_id="AAG57693.1"  
/db\_xref="GI:12516998"  
/translation="MFLHSKFDIFPRGMDLRRFITLKTVEEGSFLRASQKLCTQ  
STVFHIQOLEQERSVQLFEKIGRMCLTREGKLLPHIYELTRWMDTLREAAKESD  
PDGELRVSGETLLSYRMPQVLQRFORAPKVRLLSLOSLNCYVIRDALLNDEADVGF  
YRVGNDALNRELGEQSLVLVASPQIADVDFTEGRHNACSFITNEPQCVFRQIFES  
TLRQRRITVENTIELLSIESIKRCVANGVSYLPRFAVEKELESSELIELPFGEHSQ  
TITAMCAHHAAGRAVSPAMHTFIQCVEDCFLPG"  
5062. .5649  
/gene="yfiK"  
/note="synonym: Z3861"  
5062. .5649  
/gene="yfiK"  
/function="orf; Unknown function"  
/note="Residues 1 to 195 of 195 are 97.43 pct identical to

residues 1 to 195 of 195 from Escherichia coli K-12 Strain  
MG1655: B2578"  
/codon\_start=1  
/transl\_table=11  
/product="orf, hypothetical protein"  
/protein\_id="AAG57694.1"  
/db\_xref="GI:12516999"  
/translation="MTPTLSAFWYTYTLTAMTPGPNNILALSSATTHGFHOSTRVLA  
GMSIGFLIWLMLCAGISFSLAVIDPAVHLISWAGAVIWLAWKIATSPTKEDGIQT  
KRISFWASPALQFVNVKIILYGVTAISTFVLPQTQALSMIVGSVLLAMIGTFGNVCW  
ALAGHLFQRLFRQYGRQNLVLLLIYCAVRIFY"  
complement (5705. .6088)  
/gene="yfiD"  
/note="synonym: Z3862"  
complement (5705. .6088)  
/gene="yfiD"  
/function="putative enzyme; Energy metabolism, carbon:  
Anaerobic respiration"  
/note="Residues 1 to 127 of 127 are 100.00 pct identical  
to residues 1 to 127 of 127 from Escherichia coli K-12  
Strain MG1655: B2579"  
/codon\_start=1  
/transl\_table=11  
/product="putative formate acetyltransferase"  
/protein\_id="AAG57695.1"  
/db\_xref="GI:12517000"  
/translation="MITGIQITKAANDDLNSFWLLDSEKGEARKIVAKGYAEDEV  
AVSKLDIEYREVPEVKEVVEGGQHLNVNLRRETLIEDAVKHPEKYPQLTIRVSG  
YAVRFNSLTPEQQRDVIAFTFESL"  
6393. .7082  
/gene="ung"  
/note="synonym: Z3864"  
6393. .7082  
/gene="ung"  
/function="enzyme; Macromolecule synthesis, modification:  
DNA - replication, repair, restr./modific'n"  
/note="Residues 1 to 229 of 229 are 99.12 pct identical to  
residues 1 to 229 of 229 from Escherichia coli K-12 Strain  
MG1655: B2580"  
/codon\_start=1  
/transl\_table=11  
/product="uracil-DNA-glycosylase"  
/protein\_id="AAG57696.1"  
/db\_xref="GI:12517001"  
/translation="MANELTMDVDVLAEEKQPYFLNTLQTVASERQSGVTIYPPQKDV  
FNARFTELDVKVVLIGQDPYHGPQAHGLAFSVRPGIATPSSLNMYKELENTIPG  
FTRPNHGYLESWARQGVLLNLTVAIRAGQASHASLGEFTDKVISLINQHREGVV  
FLMGSHAQKKGAIIIDQRHHVLAKPSPSLSAHRGFGCNHFVLANQMLEQHGETPI  
DMMPVLPASEE"  
complement (7130. .8167)  
/gene="yfiF"  
/note="synonym: Z3865"  
complement (7130. .8167)  
/gene="yfiF"  
/function="orf; Unknown function"  
/note="Residues 1 to 345 of 345 are 99.42 pct identical to  
residues 1 to 345 of 345 from Escherichia coli K-12 Strain  
MG1655: B2581"  
/codon\_start=1  
/transl\_table=11  
/product="orf, hypothetical protein"  
/protein\_id="AAG57697.1"  
/db\_xref="GI:12517002"  
/translation="MNDKMGKSGKVKVMYVRSDDSDKRTNPNRTGKGGGRPGKSR  
DGGRRPARDDKOSQPRDRKWEBSPWRTVSRAPGDETPKADHGISGKSFIDPEVLR  
QRAETRVYGENACQALFQSRPEAIVRAWFIQSVTPRFKALRWMAANRKAHXHVUDEA

Query Match 91.7%; Score 687.6; DB 1; Length 12337;  
Best Local Similarity 94.8%; Pred. No. 1.5e-204;  
Matches 711; Conservative 0; Mismatches 39; Indels 0; Gaps 0;  
QY 1 GATCCATAACCCGAACCTATCGAAATATCGAATCTAGATATATAAACAATTGATTTT 60  
|||||

Db 4953 GATCCATAACCCCAACCTATCGAAATATCGAATTGAAATGTAACATTCATTTT 5012

QY 61 TTTAAATGTTCCGTGTGCGGTACTGTCTACCAAAAACAGAGAGATTAACAAGTGACCCGAC 120

Db 5013 TTTAAATATTTCTGTGTGCGGTACTGTCTACCAAAAACAGAGAGATTAACAAGTGACCCGAC 5072

QY 121 CCTTTTAAGTCTTTTGGACTTACACCCCTGATTACCGCTATGACGCCAGACCGAACAA 180

Db 5073 CTTTTTAAGTCTTTTGGACTTACACCCCTGATTACCGCTATGACGCCAGACCGAACAA 5132

QY 181 TATTCCTGCCCTTAGCTCTGCTACAGTCGCGATGATTTCGCAAAAGTACCCCGCTGTGC 240

Db 5133 TATTCCTGCCCTTAGCTCAAGCACAACGACGATTCATCAGATACCCCGCTGTGC 5192

QY 241 AGGATGAGTCTGGGATTTTGAATTGATGTTACTGTGTCGGGCGCATTTCACTTTCACT 300

Db 5193 AGGATGAGTCTGGGATTTTGAATTGATGTTACTGTGTCGGGCGCATTTCACTTTCACT 5252

QY 301 GGCAGTGTGACCCCGGACGGGTACACCTTTGAGTTGGGCGGCGGCATATATTGT 360

Db 5253 GGCTGTATCGACCCCGGACGTGTGATCTGTTGAGTTGGGCGGCGGCATATATTGT 5312

QY 361 CTGGCTGGCGTGGAAATCGCCACGACCCCAACAAGAGACGACTTCAGGCAAAACC 420

Db 5313 CTGGCTGGCGTGGAAATCGCTACAGTCCAACAAGAGACGCGCTTCAGCAAAACC 5372

QY 421 AATCAGCTTTTGGGCCAGCTTTGCTTGCAGTTTGTGAACGTCAAAATCATTTGTACGG 480

Db 5373 AATCAGTTTGGGCCAGCTTTGCTTGCAGTTTGTGAACGTCAAAATCATTTGTACGG 5432

QY 481 TGTACGGCACTGTGACGTTTGTCTGCCGCAACACAGCGGTTAAGCTGGGTAGTTGG 540

Db 5433 TGTACGGCACTGTGACGTTTGTCTGCCGCAACACAGCGGTTAAGCTGGGTAGTTGG 5492

QY 541 CGTCAGCGTTTGTGCGCATGATTGGGACGTTTGGCAATGTGTCTGGGCGCTGGCGGG 600

Db 5493 CGTCAGCGTTTGTGCGCATGATTGGGACGTTTGGCAATGTGTCTGGGCGCTGGCGGG 5552

QY 601 GCATCTGTTTACGCGATGTTTTCGCCAGTATGTCGCCAGTTAATATCTGCTTGGCCCT 660

Db 5553 GCATCTGTTTACGCGATGTTTTCGCCAGTATGTCGCCAGTTAATATCTGCTTGGCCCT 5612

QY 661 GTTGTGTTCTATTGCGCGGTACGCAATTTTCTATTAAAGAAAAAGCGGAAGGTCCG 720

Db 5613 GTTGTGTTCTATTGCGCGGTACGCAATTTTCTATTAAAGAAAAAGCGGAAGGTCCG 5672

QY 721 CCTCTCCGCTTAGTAATTGCTACTTAAG 750

Db 5673 CCTCTCCGCTTAGTAATTGCTACTTAAG 5702

RESULT 10

AP002562 270365 bp DNA linear BCT 20-MAR-2004

LOCUS AP002562 Escherichia coli O157:H7 DNA, complete genome, section 13/20.

DEFINITION AP002562 BA000007

ACCESSION AP002562.1 GI:13362858

VERSION

KEYWORDS

SOURCE Escherichia coli O157:H7

ORGANISM Escherichia coli O157:H7

REFERENCE 1 Enterobacteriaceae; Escherichia.

AUTHORS Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S., Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T., Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T., Sasaki,K., and Shinagawa,H.

TITLE Complete nucleotide sequence of the prophage VT2-Sakai carrying the verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7 derived from the Sakai outbreak

JOURNAL Genet. Syst. 74 (5), 227-239 (1999)

MEDLINE 20198780

PUBMED 10734605

REFERENCE 2

AUTHORS Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M., Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and Hayashi,T.

TITLE Comparative analysis of the whole set of rRNA operons between an enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an Escherichia coli K-12 strain MG1655

JOURNAL Syst. Appl. Microbiol. 23 (3), 315-324 (2000)

MEDLINE 20557356

PUBMED 11108008

REFERENCE 3

AUTHORS Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S., Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T., Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasaki,K., and Shinagawa,H.

TITLE Complete nucleotide sequence of the prophage VT1-Sakai carrying the Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli O157:H7 strain derived from the Sakai outbreak

JOURNAL Gene 258 (1-2), 127-139 (2000)

MEDLINE 20564182

PUBMED 11111050

REFERENCE 4

AUTHORS Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K., Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T., Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasaki,K., Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and Shinagawa,H.

TITLE Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12

JOURNAL DNA Res. 8 (1), 11-22 (2001)

MEDLINE 21156231

PUBMED 11258796

REFERENCE 5

AUTHORS (bases 1 to 270365) Hattori,M., Ishii,K. and Shiba,T.

TITLE Direct Submission

JOURNAL Submitted (26-JUN-2000) Masahira Hattori, Kitasato Institute for Life Sciences, Kitasato University; Kitasato 1-15-1, Sagami-hara, Kanagawa 228-8555, Japan

COMMENT (E-mail:hattori@genome.ls.kitasato-u.ac.jp, URL:http://genome.ls.kitasato-u.ac.jp/, Tel:81-42-778-8194, Fax:81-42-778-8193)

genome project

This work was done in collaboration with Tetsuya Hayashi, Makoto Ohnishi, Keisuke Nakayama (Miyazaki Medical College), Kozo Makino, Ken Kurokawa, Katsushi Yokoyama, Masashi Tanaka, Takeshi Honda, Teruo Yasunaga, Hideo Shinagawa (Osaka University), Takahiro Murata (Shinsu University), Chang-Gyun Han, Eiichi Ohtsubo, Toru Tobe, Chihiro Sasaki (University of Tokyo), Hideto Takami (Japan Marine Science and Technology Center), Naotake Ogasawara (Nara Institute of Science and Technology), Satoru Kuhara (Kyushu University), and supported by the Research for the Future Program of the Japan Society for the Promotion of Science.

FEATURES

source

1. 270365

/organism="Escherichia coli O157:H7"

/mol\_type="genomic DNA"

/strain="O157:H7"

/sub\_strain="RIMD 0509952"

/db\_xref="taxon:83334"

/complement(435..1220)

/gene="EC63388"

/complement(435..1220)

/gene="EC63388"

/note="similar to SEEB\_ECOLI gi|1788871 percent identity 98 in 261 aa (Conserved in E.coli K-12)"

/codon\_start=1

/evidence=not\_experimental

/transl\_table=11

/product="enhanced serine sensitivity"

/protein\_id="BAB36811.1"

/db\_xref="GI:13362859"

/translation="MTLMSETKNELEDLLEKATEPAHRAFPRTLLSESTVWVPGTAA

OGEAVVEDSALDLQHWKEDGTSVIPFTSLALQAVDEQAFFVMPVRTLLFEMTLG



ETLFLNAKLP TGKEFMPREISLIGEGNPLSSOEILEGGEISLISEVAEPPAQMIDS  
LTTLEFKTIKVPKRAFICISIKENEAOPLLIGIEADGDIEEIIQVAGSVATDTLPGDE  
PIDICQVKGEKISHFITEHIAPIFYERRWGGFLRDFKQNRIT"  
gene  
complement (1354. .2724)  
/gene="ECs3389"  
complement (1354. .2724)  
/note="similar to PEPB\_ECOLI gi|1788872 percent identity  
99 in 456 aa (Conserved in E.coli K-12)"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="putative peptidase"  
/protein\_id="BAB36812.1"  
/db\_xref="GI:13362860"  
/translation="MPSPCNANHLAEKSWGCCQLALLQRRITKMTKTEAMKITLSTQPAD  
ARWGEKATYSINNDGITLHLNGADDLGLQRAARKIDGLIKHVQLSGEWDADRCWA  
FMQGYKAPKGTARKVEMPDLDDAQOQELDNRLMIIDWVRTINAPAEIGPSQLQRAV  
DLISNVASDRVYRITKGEDLREOGYMGHTVGRGSESPVLIALIDYNPTGDKAPVY  
ACLVGKGITFDSSGYSIKQTAFMDSKMSDMGGAATVTGALAFATIRGLNKRVKLFLCC  
ADNLISGNAFKLDIITYRNGKVEVMTDAEGLVLADGLIDASAQKPEMIIIDAAITL  
TGAAKTALGNDYHALFSFDLALAGRLASASQENEPWRLLPLAEFHRSQLPSNFAELN  
NTGSAAYPAGASTAAGFLSHFVENYQQGWLHIDCSATYRKAPVEQWSAGATGLGVRTI  
ANLITA"  
gene  
complement (2815. .3015)  
/gene="ECs3390"  
complement (2815. .3015)  
/gene="ECs3390"  
/note="identical to YFHJ\_ECOLI gi|1788873 (Conserved in  
E.coli K-12)"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="BAB36813.1"  
/db\_xref="GI:13362861"  
/translation="MGLKMTDSREIGEALYDAYPDLDPKTVRFTDMHQWICDLED added  
DPQASNEKILEAILLVWLDEAE"  
gene  
complement (3027. .3362)  
/gene="ECs3391"  
complement (3027. .3362)  
/gene="ECs3391"  
/note="similar to FER\_ECOLI gi|119999|sp|P25528 percent  
identity 100 in 111 aa (Conserved in E.coli K-12)"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="[2FE-2S] ferredoxin"  
/protein\_id="BAB36814.1"  
/db\_xref="GI:13362862"  
/translation="MPKIVILPHQDLCPDGAVLEANSGETILDAALRNGIEIEHACEK  
SCACTCHCTVREGFDSLPESSSEQEDMDLKAWGLEPESRLSCQARTDDEDLVEIIPR  
YTINHAREH"  
gene  
complement (3364. .5214)  
/gene="ECs3392"  
complement (3364. .5214)  
/gene="ECs3392"  
/note="similar to HSCA\_ECOLI gi|1788875 percent identity  
100 in 616 aa (Conserved in E.coli K-12)"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="molecular chaperone"  
/protein\_id="BAB36815.1"  
/db\_xref="GI:13362863"  
/translation="MALLQISEPGLSAPHQRRLAAGIDLTGTSNLVATVRSQAETL  
ADHEGRHLPSVHYOOQGHVSVDARTNALDTANTISSVKRLMGRSLADIQORYPH  
LPYQOASENGLEPMETAGLNLPRVSADILKALARAATEAGVAVITVPAYF  
DDAOROGTKDAARLAGLHVLRLINEPTAAIAYGDSGOEVIAYVYDLGGGTFDISIL  
RUSRGVFEVLATGGDSALGGDDFDHLADYIREQAGIPDRSDNRVQRELDAAIYAKI  
ALSDADSVTVNAGMOGEISRQENELIAPLVKRTLLACRALXDAAGVADDEVLEVM  
VGGSTRVPLVREVRGEFFGRPLTISIDPKVVAIGAAIQADILVGNKPBSEMLLDVI  
PLSLGLETMGVLEKVIIPRNTTIPVARAQDFTFKDGTAMSIHVMQGERELVQDCRS

LARFALRGIPALPAGAHIRVTFQVDADGLLSVTAMEKSTGVEASIQVKPSYGLTDE  
IASMIKDSMSYAEOQVRAKMLAEQKVEARVLESILHGAALADALLSAERQV1DDAA  
AHISEVAQGDVDVAIEQAIKNVDKOTQDFAAARMQSVRALKGHSVDEV"  
gene  
complement (5231. .5746)  
/gene="ECs3393"  
complement (5231. .5746)  
/note="similar to HSCB\_ECOLI gi|1788876 percent identity  
100 in 171 aa (Conserved in E.coli K-12)"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="molecular chaperone"  
/protein\_id="BAB36816.1"  
/db\_xref="GI:13362864"  
/translation="MDYFTLFLGLPARYQLDTOALSFRQDLQROYHPDKFAGSQAEQ  
LAAVQSATINQAWQTLRHLMLRAEYLLSHGFDLASEQHTVRDTAFMLEQLEBEL  
DEIEQAKDEARLESFIKRVKMFDRHQLMVEQDLNBTWDAADTVRKLRFLDKLRSS  
AEQLEEKLLDF"  
gene  
complement (5842. .6165)  
/gene="ECs3394"  
complement (5842. .6165)  
/gene="ECs3394"  
/note="similar to ISCA\_ECOLI gi|1788877 percent identity  
100 in 107 aa (Conserved in E.coli K-12)"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="putative iron-binding protein"  
/protein\_id="BAB36817.1"  
/db\_xref="GI:13362865"  
/translation="MSITLSDSAARVNTFLANRGKFGRLGVRTSGSGMAVYLEF  
VDEPTPEDIVFEDKGVKVVVDGKSLQFLDGTQLDFVKEGLNEGFKFTNPVNDKCGCG  
ESFHY"  
Query Match 91.7%; Score 687.6; DB 1; Length 270365;  
Best Local Similarity 94.8%; Pred. No. 2.9e-204;  
Matches 711; Conservative 0; Mismatches 39; Indels 0; Gaps 0;  
gene  
CDS  
1 GATCCATAACCCCAAAACCTATCGAAATATCGAATCTAGAATATATAAAACATTCTTTT 60  
61135 GATCCATAACCCCAAAACCTATCGAAATATCGAATTTAGAAATGTAAACATTCATTTT 61194  
61 TTAATGTTCCGTGTCGGGTACTGTCTACCAAAACAGAGAGATTAACAAGTGACACCGAC 120  
61195 TTAATATCTGTGTGGGTACTGTCTACCAAAACAGAGAGATTAACAAGTGACACCGAC 61254  
121 CCTTTAAGTCTTTTGGACTTACACCCCTGATTAACCGCTATGACGCCAGCAACAA 180  
61255 CCTTTAAGTCTTTTGGACTTACACCCCTGATTAACCGCATGACGCCAGCAACAA 61314  
181 TATTCGCGCCCTAGCTCTGCTACGTCGCAAGGATTTTCGCAAAAGTACCCGCGTGGC 240  
61315 TATTCGCGCCCTAGCTCAGCCACACGACGAGATTTCATCAGAGTACCCGCGTGGC 61374  
241 AGGATGAGTCGGGATTTTGTGATGTGATGTTACTGTGTGCGGGCAATTCATTTTCACT 300  
61375 AGGATGAGTCGGGATTTTATCGTCACTGTGCTGTGTGCGGGCAATTCATTTTCACT 61434  
301 GGCAGTGATTGACCCGGCAGCGGTACACCTTTTGAAGTTGGCGGGGGCGCATATATTGT 360  
61435 GGCTGTTATCGACCCGGCAGCGGTGATCTGTGAGTTGGCGGGGGCGCATATATTGT 61494  
361 CTGGCTGGCGTGAATAATCGCCACGACCCCAACAAAGAGACGGACTTCAGGCCAAAACC 420  
61495 CTGGCTGGCGTGAATAATCGCTACCAAGTCCCAACAAAGAGACGGCTTCAGACAAAACC 61554  
421 AATCAGCTTTTGGCGCAGCTTTGCTTGTGACGTTGTGAACGTCAAAATCATTTTGTACGG 480  
61555 AATCAGCTTTTGGCGCAGCTTTGCTTGTGACGTTGTGAACGTCAAAATCATTTTGTACGG 61614  
481 TGTTACGGCACTGTGACGTTTGTGCTGCGGCAACACAGCGCTTAAGCTGGGTAGTTGG 540  
|||||



Db 61615 TGTTACGGCACTGTGCACATTGTCTCCGCCAACAACAGCGCTTAAGCTGATAGTTGG 61674

QY 541 CGTCAGCGTTTGTGCTGGCGATGATGGAGCGTTTGGCAATGTGTGCTGGCGCTGGCGGG 600

Db 61675 CGTCAGCGTTTGTGCTGGCAATGATGGAGCGTTTGGTAATGTGTGCTGGCGCTGGCGGG 61734

QY 601 GCATCTGTTTACGCGATGTTTCCGACGATGATGTCGCCAGTTAAATATCGTCTGGCCCT 660

Db 61735 GCATCTGTTTACGCGATGTTTCCGACGATGATGTCGCCAGTTAAATATCGTCTGGCGCT 61794

QY 661 GTTGCTGCTATATGCGCGGTAAGCATTTTCTATTAAAGAAAAAGCGGAAGAGGTGCG 720

Db 61795 GCTGCTGATCTATTGCGCGGTAAGCATTTTCTATTAAAGAAAAAGCGGAAGAGGTGCG 61854

QY 721 CCTCTCCGCTTAGTAAGTTGCTACTTAAG 750

Db 61855 CCTCTCCGCTTAGTAAGTTGCTACTTAAG 61884

RESULT 11

AC126452 194171 bp DNA linear HTG 15-SEP-2004

LOCUS Mus musculus chromosome UNKNOWN clone RP23-215C15, WORKING DRAFT

DEFINITION AC126452

SEQUENCE, 7 unordered pieces.

AC126452

AC126452.3 GI:52077982

HTG: HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ACTIVEFIN.

KEYWORDS Mus musculus (house mouse)

SOURCE Mus musculus

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 194171)

AUTHORS Wilson,R.K.

TITLE The sequence of Mus musculus clone

JOURNAL Unpublished

AUTHORS 2 (bases 1 to 194171)

McPherson,J.D. and Waterston,R.H.

TITLE Direct Submission

JOURNAL Submitted (05-JUL-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE 3 (bases 1 to 194171)

AUTHORS Wilson,R.K.

TITLE Direct Submission

JOURNAL Submitted (15-SEP-2004) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

COMMENT On Sep 15, 2004 this sequence version replaced gi:38564407.

----- Genome Center -----

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: http://genome.wustl.edu

Contact: submissions@watson.wustl.edu

----- Project Information -----

Center project name: M\_BA0215C15

----- Summary Statistics -----

Sequencing vector: M13; 0%

Sequencing vector: plasmid; 100%

Chemistry: Dye-primer ET; 0% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 191627 bases at least Q40

Consensus quality: 191807 bases at least Q30

Consensus quality: 191942 bases at least Q20

-----

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 7 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

\* 1 1232: contig of 1232 bp in length

\* 1233 1332: gap of unknown length

\* 1333 3669: contig of 2337 bp in length

\* 3670 3769: gap of unknown length

\* 3770 9983: contig of 6214 bp in length

\* 9984 10083: gap of unknown length

\* 10084 31227: contig of 21144 bp in length

\* 31228 31327: gap of unknown length

\* 31328 69948: contig of 38621 bp in length

\* 69949 70048: gap of unknown length

\* 70049 192986: contig of 122938 bp in length

\* 192987 193086: gap of unknown length

\* 193087 194171: contig of 1085 bp in length.

FEATURES

Location/Qualifiers

1. 194171

source

/organism="Mus musculus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10090"

/chromosome="UNKNOWN"

/clone="RP23-215C15"

1. 1232

misc\_feature

/note="assembly\_name:Contig42"

1333. 3669

misc\_feature

/note="assembly\_name:Contig43"

3770. 9983

misc\_feature

/note="assembly\_name:Contig44"

10084. 31227

misc\_feature

/note="assembly\_name:Contig45"

31328. 69948

misc\_feature

/note="assembly\_name:Contig46"

70049. 192986

misc\_feature

/note="assembly\_name:Contig47

clone end:SP6

vector\_side:left"

193087. 194171

misc\_feature

/note="assembly\_name:Contig7"

ORIGIN

Query Match 83.1%; Score 623.4; DB 2; Length 194171;

Best Local Similarity 99.7%; Pred. No. 4.6e-184;

Matches 635; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 115 ACCGACCCCTTTAAGTCTTTTGGACTTACACCCCTGATTACCGCTATGACGCCAGGACC 174

Db 193122 ATGACCCCTTTTAAAGTCTTTTGGACTTACACCCCTGATTACCGCTATGACGCCAGGACC 193181

QY 175 GAACAATATTCTGCCCTTAGCTCTGCTACGTCGATGATGATTTCGTCAAAGTACCCGCGT 234

Db 193182 GAACAATATTCTGCCCTTAGCTCTGCTACGTCGATGATGATTTCGTCAAAGTACCCGCGT 193241

QY 235 GCTGGCAGGATGAGTCTGGGATTTTGAATGTGATGTTACTGTGTGGCGCATTTTCATT 294

Db 193242 GCTGGCAGGATGAGTCTGGGATTTTGAATGTGATGTTACTGTGTGGCGCATTTTCATT 193301

QY 295 TTCACTGCGAGTGAATTGACCCGCGACGCGTACACCTTTGAGTTGGCGGGCGGCATA 354

Db 193302 TTCACTGCGAGTGAATTGACCCGCGACGCGTACACCTTTGAGTTGGCGGGCGGCATA 193361

QY 355 TATTGTCTGGCTGGCGGTGAAATCGCCACCAAGCCCAACAAGGAAGACGGAATTCAAGC 414

Db 193362 TATTGTCTGGCTGGCGGTGAAATCGCCACCAAGCCCAACAAGGAAGACGGAATTCAAGC 193421

QY 415 AAAACCAATCAGCTTTTGGGCGCAGCTTTGCTTGCAGTTGTGTAACGTCAAATCATTTT 474

Db 193422 AAAACCAATCAGCTTTTGGGCGCAGCTTTGCTTGCAGTTGTGTAACGTCAAATCATTTT 193481

QY 475 GTACGGTGTAAACGCACTGTGCAAGTTGTCTGCCGAAACACAGCGCTTAAGCTGGGT 534

Db 193482 GTACGGTGTAAACGCACTGTGCAAGTTGTCTGCCGAAACACAGCGCTTAAGCTGGGT 193541

QY 535 AGTTGCGGTACGCTTTTGTCTGCGATGATTGGAGCGTTTGGCAATGTGTCTGGCGGCT 594

Db 193542 AGTTGCGGTACGCTTTTGTCTGCGATGATTGGAGCGTTTGGCAATGTGTCTGGCGGCT 193601

QY	595	GGCGGGGCATCTGTTTTCAGCGATTGTTTCGCCAGTATGGTCGCCAGTTAAATATCGTCT	654
Db	193602	GGCGGGGCATCTGTTTTCAGCGATTGTTTCGCCAGTATGGTCGCCAGTTAAATATCGTCT	1936
QY	655	TGCCCTGTGCTGCTGTTCTATTGCGCGGTACGCATTTCTATTAAACGAAAAAGCGGAAAGA	714
Db	193662	TGCCCTGTGCTGCTGTTCTATTGCGCGGTACGCATTTCTATTAAACGAAAAAGCGGAAAGA	1937
QY	715	GGTCGCCCTCTCC-GCTTAGTAACTTGCTACTTAAG	750
Db	193722	GGTCGCCCTCTCTCCGCTTAGTAACTTGCTACTTAAG	193758
RESULT 12			
E49391			
LOCUS	E49391	588 bp	DNA
DEFINITION	Process for producing L-amino acid.		linear
ACCESSION	E49391		PAT 31-JAN-2002
VERSION	E49391.1	GI:18628082	
KEYWORDS	JP 2000189180-A/11.		
SOURCE	Escherichia coli		
ORGANISM	Escherichia coli		
REFERENCE	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.		
AUTHORS	1 (bases 1 to 588)		
TITLE	Rivishitsu,V.A., Zakataeva,N.P., Nakanishi,K., Aryoshin,V.V.,		
JOURNAL	Toroshin,P.V. and Tokumakova,I.R.		
	Process for producing L-amino acid		
	Patent: JP 2000189180-A 11 11-JUL-2000;		
	AJINOMOTO CO INC		
COMMENT	OS Escherichia coli		
	PN JP 2000189180-A/11		
	PD 11-JUL-2000		
	PF 28-DEC-1999 JP 1999373651		
	PR 30-DEC-1998 RU 98124016,09-MAR-1999 RU 99104431 PI		
	VITARI ARUKAJEVICHI RIVISHITSU,NATARIYA PAVUROVUNA		
	ZAKATAEVA,		
	PI KAZUO NAKANISHI,VLADIMIR VENIYAMINOVICHI ARYOSHIN, PI PETER		
	VIRAJIMIROVICHI TOROSHIN,IRINA RIVOYUNA TOKUMAKOVA PC		
	C12N15/09,C12N1/21,C12P13/04//((C12N1/21,C12R1:19),(C12P13/04,PC		
	C12R1:19),		
	PC C12N15/00		
CC			
FH			
FT			
FEATURES			
source			
	1..588	location/Qualifiers	
	/organism="Escherichia coli"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:562"		
ORIGIN			
Query Match	78.4%;	Score 588;	DB 6;
Best Local Similarity	100.0%;	Pred. No. 1.8e-173;	Length 588;
Matches 588;	Conservative 0;	Mismatches 0;	Indels 0;
			Gaps 0;
QY	110	GTGACACCGGACCCCTTTTAAGTCCTTTTGGACTTACACCGTGATTACCGCTATGACGCCA	169
Db	1	GTGACACCGGACCCCTTTTAAGTCCTTTTGGACTTACACCGTGATTACCGCTATGACGCCA	60
QY	170	GGAACGAAACATATTCTCGCCCTTAGCTCTGCTACGTGCGATGGAATTCGTCAAAGTACC	229
Db	61	GGAACGAAACATATTCTCGCCCTTAGCTCTGCTACGTGCGATGGAATTCGTCAAAGTACC	120
QY	230	CGCGTGTGGCAGGATGAGTCTGGGATTTTGTATTGTGATGTTACTGTGTGCGGGCAT	289
Db	121	CGCGTGTGGCAGGATGAGTCTGGGATTTTGTATTGTGATGTTACTGTGTGCGGGCAT	180
QY	290	TCAATTTTCACTGGCAGTGATTGACCCGGCAGCGGTACACCTTTTGAAGTTGGCGGGGCG	349
Db	181	TCAATTTTCACTGGCAGTGATTGACCCGGCAGCGGTACACCTTTTGAAGTTGGCGGGGCG	240

QY	350	GCATATATTGCTGCGCTGGCGGTGAAAAATCGCCACAGCCCAACAAAGGACGGACTT	409		
Db	241	GCATATATTGCTGCGCTGGCGGTGAAAAATCGCCACAGCCCAACAAAGGACGGACTT	300		
QY	410	CAGGCAAAACCAATCAGCTTTTGGGCCAGCTTTGCTTGCAGTTTGTGAACGTCAAAATC	469		
Db	301	CAGGCAAAACCAATCAGCTTTTGGGCCAGCTTTGCTTGCAGTTTGTGAACGTCAAAATC	360		
QY	470	ATTTTGTACGGTGTACGGCACTGTCAGCGTTTGTCTGCGCAACACAGCGTTAAAGC	529		
Db	361	ATTTTGTACGGTGTACGGCACTGTCAGCGTTTGTCTGCGCAACACAGCGTTAAAGC	420		
QY	530	TGGGTAGTTGGCGTCAAGCTTTTGTCTGCGCATATTGGGACGTTTGCAATGTGCTGG	589		
Db	421	TGGGTAGTTGGCGGTCAAGCTTTTGTCTGCGCATATTGGGACGTTTGCAATGTGCTGG	480		
QY	590	GCGCTGGCGGGGCATCTGTTTCAAGCGATTGTTTCGCCAGTATGCTCGCCAATTAAATATC	649		
Db	481	GCGCTGGCGGGGCATCTGTTTCAAGCGATTGTTTCGCCAGTATGCTCGCCAATTAAATATC	540		
QY	650	GTGCTTGCCCTGTGCTGCTGCTATTGCGCGGTACGCAATTTCTATTAA	697		
Db	541	GTGCTTGCCCTGTGCTGCTGCTATTGCGCGGTACGCAATTTCTATTAA	588		
RESULT 13					
AX030083		588 bp	DNA	linear	PAT 16-SEP-2000
LOCUS	AX030083				
DEFINITION	Sequence 13 from Patent EP1016710.				
ACCESSION	AX030083				
VERSION	AX030083.1	GI:10190300			
KEYWORDS	.				
SOURCE	Escherichia coli				
ORGANISM	Escherichia coli				
REFERENCE	1				
AUTHORS	Nakanishi, K., Aleshin, V.V., Livshits, V.A., Tokmakova, I.L., Troshin, P.V. and Zakataeva, N.P.				
TITLE	Method for producing l-amino acids				
JOURNAL	Patent: EP 1016710-A 13 05-JUL-2000;				
	AJINOMOTO KK (JP)				
FEATURES	location/Qualifiers				
source	1..588				
	/organism="Escherichia coli"				
	/mol_type="unassigned DNA"				
	/db_xref="taxon:562"				
CDS	1..588				
	/note="unnamed protein product"				
	/codon_start=1				
	/transl_table=11				
	/protein_id="CAC09316.1"				
	/db_xref="GI:10190301"				
	/translation="MTPTLISAFWYTLITAMTPGPNNILALSSATSHGFROSTRVLA				
	GMSLGLIIVMLCAGISFSLAVIDPAVHLMSWAGAAIYIVLWAKIATSPTEDEGLQAP				
	KPISFWASPALQFVNKILILYGTALSTFVLPEQTQALSWVGVSVLLAMIGTFGNVCWA				
	ALAGHLFQRLFRQYGRQLNIVLALLLVYCAVRIFY"				
ORIGIN					
	Query Match	78.4%;	Score 588;	DB 6;	Length 588;
	Best Local Similarity	100.0%;	Pred. No. 1.8e-173;		
	Matches 588;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	110	GTGACACCCGACCCCTTTTAAGTCTTTTGGACTTACACCCCTGATTACCGCTATGACGCCA	169		
Db	1	GTGACACCCGACCCCTTTTAAGTCTTTTGGACTTACACCCCTGATTACCGCTATGACGCCA	60		
QY	170	GGACCCGAACAATATTCTCGCCCTTAGCTCTGCTACGTCGATGGATTTCGTCAAAGTACC	229		
Db	61	GGACCCGAACAATATTCTCGCCCTTAGCTCTGCTACGTCGATGGATTTCGTCAAAGTACC	120		
QY	230	CGCGTGCTGGCAGGATGAGTCTGGGATTTTGGATTGTGATGTACTGTGTGCGGGCAAT	289		

Db 121 CGCGTGTGGCAGGAGTAGTCTGGGATTTTGTATGTGATGTTACTGTGTGCGGCATT 180

QY 290 TCATTTTCACTGGCAGTGAATTGACCCGGCAGCGGTACACTTTTGAAGTTGGCGGGCG 349

Db 181 TCATTTTCACTGGCAGTGAATTGACCCGGCAGCGGTACACTTTTGAAGTTGGCGGGCG 240

QY 350 GCATATATTTGTCTGGCTGGCGTGGAAATCGCCACACGCCAACAAAGAGAGACGACTT 409

Db 241 GCATATATTTGTCTGGCTGGCGTGGAAATCGCCACACGCCAACAAAGAGAGACGACTT 300

QY 410 CAGGCAAAACCAATCAGCTTTTGGGCCAGCTTTGCTTTGCAGTTGTGAACGTCAAATC 469

Db 301 CAGGCAAAACCAATCAGCTTTTGGGCCAGCTTTGCTTTGCAGTTGTGAACGTCAAATC 360

QY 470 ATTTGTACGGTGTACGGCACTGTGCAGCTTTGTTCTGCCGCAACACAGGCGTTAAGC 529

Db 361 ATTTGTACGGTGTACGGCACTGTGCAGCTTTGTTCTGCCGCAACACAGGCGTTAAGC 420

QY 530 TGGGTAGTTGGCGTACGCGTTTGTCTGGCGATGATGGGACGTTTGGCAATGTGCTGG 589

Db 421 TGGGTAGTTGGCGTACGCGTTTGTCTGGCGATGATGGGACGTTTGGCAATGTGCTGG 480

QY 590 GCGCTGGCGGGCATCTTTTACGCGATTTGTTTCCGAGTATGTCGCCAGTTAATATC 649

Db 481 GCGCTGGCGGGCATCTTTTACGCGATTTGTTTCCGAGTATGTCGCCAGTTAATATC 540

QY 650 GTGCTTGCCCTGTGCTGTCTATTGCGCGGTACGCATTTTCTATTAA 697

Db 541 GTGCTTGCCCTGTGCTGTCTATTGCGCGGTACGCATTTTCTATTAA 588

RESULT 14

AL627275

LOCUS AL627275 145050 bp DNA linear BCT 04-JUL-2003  
DEFINITION Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18, complete chromosome, segment 11/20.

ACCESSION

AL627275 AL513382

VERSION

AL627275.1 GI:16503698

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

1 (bases 1 to 145050)

Parkhill,J., Dougan,G., James,K.D., Thomson,N.R., Pickard,D.,

Wain,J., Churcher,C., Mungall,K.L., Bentley,S.D., Holden,M.T.G.,

Sebahia,M., Baker,S., Basham,D., Brooks,K., Chillingworth,T.,

Conerton,P., Cronin,A., Davis,P., Davies,R.M., Dowd,L., White,N.,

Farrar,J., Feltwell,T., Hamlin,N., Hague,A., Hien,T.T., Holroyd,S.,

Jagels,K., Krogh,A., Larsen,T.S., Leather,S., Moule,S., O'Gaora,P.,

Parry,C., Quail,M., Rutherford,K., Simmonds,M., Skelton,J.,

Stevens,K., Whitehead,S. and Barrell,B.G.

Complete genome sequence of a multiple drug resistant Salmonella

enterica serovar Typhi CT18

JOURNAL Nature 413 (6858), 848-852 (2001)

MEDLINE 21534947

PUBMED 11677608

REFERENCE 2 (bases 1 to 145050)

AUTHORS Parkhill,J.

JOURNAL Direct Submission

Submitted (25-OCT-2001) Submitted on behalf of the Salmonella

sequencing team, Sanger Centre, Wellcome Trust Genome Campus,

Hinxton, Cambridge CB10 1SA, UK

E-mail: parkhill@sanger.ac.uk

COMMENT

Notes:

Details of S. typhi sequencing at the Sanger Centre are available

on the World Wide Web.

(URL, http://www.sanger.ac.uk/Projects/S\_typhi/).

FEATURES

source

1. .145050  
/organism="Salmonella enterica subsp. enterica serovar Typhi"  
/mol\_type="genomic DNA"

/strain="CT18"  
/db\_xref="taxon:90370"  
636. .640  
/note="possible RBS"  
647. .1003  
/gene="STY2720"  
/note="synonym: yffb"  
647. .1003  
/gene="STY2720"  
/note="Orthologue of E. coli yffb (YFPB\_ECOLI); Fasta hit to YFPB\_ECOLI (118 aa), 81% identity in 118 aa overlap"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="CAD02682.1"  
/db\_xref="GI:16503699"  
/db\_xref="GOA:Q8Z4S3"  
/db\_xref="UniProt/TREMBL:Q8Z4S3"  
/translation="MTLYGIRKCDTIKARRMLBEHGIDYRPHDYRVGIDLP.LLNT FIALGNQPLINTRGTTWRKLDEGHRSGITDADSAAALMVEQPAIKRPLLCAPKPM LIGFSERYQQFPEDEV"  
995. .1000  
/gene="STY2720"  
/note="possible RBS"  
1007. .2134  
/gene="STY2721"  
/note="synonym: dape"  
1007. .2134  
/gene="STY2721"  
/note="Orthologue of E. coli dape (DAPE\_ECOLI); Fasta hit to DAPE\_ECOLI (375 aa), 94% identity in 375 aa overlap"  
/codon\_start=1  
/transl\_table=11  
/product="succinyl-diaminopimelate desuccinylase"  
/protein\_id="CAD02683.1"  
/db\_xref="GI:16503700"  
/db\_xref="GOA:Q8Z4S2"  
/db\_xref="UniProt/TREMBL:Q8Z4S2"  
/translation="MSCPVELTQOLIRPSLSPPDVGCQALMERLRKIGFTIEHMD FGDTONFWARGETLAFAGHTDVVPAGDVRWINPPEPTLRDMLFGRGADMKG SLAMVVAERFVAQHPHRGR.LAPLITSDEASAKNGTVVVEALMARNERLDYCLV GEPSTFIVDVKNRGRLTICNLTIHGVQHVAYPHLADNPVRAAPFLNELVAIE WDRGNDFPATSMQIANIQAGTSNNVIEGELFVQFNFRFSTELTDEMIKERVHALL KHOLRTYVDWMLSGOPFLTARGKLDVAVVNAIEHYNBIKPOLITGTSDSGRFIARMG AQVVEIGPVNATIHKINECTVNADYLQLARMYQRI MEQLVA"  
1031. .1978  
/gene="STY2721"  
/note="Pfam match to entry PF01546 Peptidase\_M20, Peptidase family M20/M25/M40, score 315.30, E-value 7.4e-91"  
1187. .1216  
/gene="STY2721"  
/notes="PS00758 ArgE / dape / ACY1 / CPG2 / yscs family signature 1"  
1295. .1414  
/gene="STY2721"  
/note="PS00759 ArgE / dape / ACY1 / CPG2 / yscs family signature 2"  
2150. .2153  
/note="possible RBS"  
2162. .2362  
/gene="STY2722"  
2162. .2362  
/gene="STY2722"  
/note="STY2722"  
/note="Similar to Escherichia coli similar to TR:P77259 (EMBL:D90875) (66 aa) fasta scores: E(): 1e-27, 95.5% id in 66 aa. Contains hydrophobic, possible membrane-spanning region"  
/codon\_start=1  
/transl\_table=11  
/product="putative membrane protein"  
/protein\_id="CAD02684.1"  
/db\_xref="GI:16503701"



/db\_xref="UniProt/TREMBL:Q8XF02"
 /translation="MDWLAKYMWILVFLVGLLVNIKOLKRIDHKKFLANKPELPP
 HRDFNDKMDDEDWPKDQPKK"
 complement(2389..4404)
 /gene="STY2723"
 complement(2389..4404)
 /gene="STY2723"
 /note="Orthologue of E. coli YPFI\_ECOLI; Fasta hit to
 YPFI\_ECOLI (671 aa), 72% identity in 662 aa overlap"
 /codon\_start=1
 /transl\_table=11
 /product="conserved hypothetical protein"
 /protein\_id="CAD02685.1"
 /db\_xref="GI:16503702"
 /db\_xref="GOA:Q8Z4S1"
 /db\_xref="UniProt/TREMBL:Q8Z4S1"
 /db\_xref="UniProt/TREMBL:Q8Z4S1"
 /translation="MSDIDALRALTSQMTQEGIRRLIVISGDAACGERAEAIRAVLP
 GDWVAPDAPAOPRCTPQALQTLGREFRHAIFDAMQGFDAAPALSGTLQAGSML
 LIMPYEETWESRPDTSLRWSDCAPIPTPOFAOHLKRTLSRDPOTLLMROROPFCWP
 SYPSREWRPATGEPEQOAAILSLREMPGVATVIAPRGKKSALAGQFISOMAGT
 AIVTAPAKTADILIAFAGERFCFMAPDALISGARADWLIVDEAAIIPALLIOLVS
 RPEIRLLTTVOEGYEGTGRGFLKFCARFQHLRFTLRQPRWAPCEPLENIVSEALI
 FDDEAFQAAPHGIAISAFYQAWGKTPALPRAVYQLSGAHYRTSPDLRRMDAPG
 QHFLQATANNRVAGALWVEEGLSAELSAQVWCGFRRPRGNLVAOSLAHGSDDLAA
 TLVGRVSRITAVHPARQREGIGQQLACACEQAQCDYLSVSGYTPBELMRFWQRCGF
 VLVRMGSHREASSGCTYAMALLPLSDAQRLAQOEHRLLRDADILITQWNGEAIPLAA
 LREQALNDEDMRELVGFAFAHRPLTSLGCLHRLLOYALPLPALGRLEKASDAEL
 CARLISGRKALLALQRAQAQALIALDAGRTQRLRDVMPGGGDHAG"
 complement(2878..3207)
 /gene="STY2723"
 /note="Pfam match to entry PF00583 Acetyltransf,
 Acetyltransferase (GNAT) family, score 32.50, E-value
 9.7e-06"
 complement(3790..3813)
 /gene="STY2723"
 /note="PS00017 ATP/GTP-binding site motif A (P-loop)"
 complement(4420..5283)
 /gene="STY2724"
 complement(4420..5283)
 /gene="STY2724"
 /note="Orthologue of E. coli P76563; Fasta hit to P76563
 (287 aa), 92% identity in 286 aa overlap. Contains
 hydrophobic, possible membrane-spanning region"
 /codon\_start=1
 /transl\_table=11
 /product="putative membrane protein"
 /protein\_id="CAD02686.1"
 /db\_xref="GI:16503703"
 /db\_xref="GOA:Q8Z4S0"
 /db\_xref="UniProt/TREMBL:Q8Z4S0"
 /db\_xref="UniProt/TREMBL:Q8Z4S0"
 /translation="MRQGRRESNNVEDRRNRPGGSLGGPGRFLPRKGIIILLVV
 VVAGYGVDLTGLLTGQPVSQQSTRSISPNDEAKFTSVILATEDPTWGLQFKMG
 RGYQPKLVMYRGMTRTGCAGQSVGPFYCPADGTVIIDLSFYDVMKNLGAADGDF
 QGYVIAHEVGHVQKLLIGIEPKVRQLQNASQTEVNRLSVRMELQADCFAGVGHSMQ
 QOGLAEAGDLBEALNAAQAIIGDRLQQQGGRVVPDSFTHTGTSQRYSWFKRGFDSGD
 PAQCNTPGKNF"
 complement(5287..5292)
 /note="possible RBS"
 complement(5415..6128)
 /gene="STY2725"
 /note="synonym: purC"
 complement(5415..6128)
 /gene="STY2725"
 /note="Orthologue of E. coli purC (PUR7\_ECOLI); Fasta hit
 to PUR7\_ECOLI (237 aa), 95% identity in 237 aa overlap"
 /codon\_start=1
 /transl\_table=11
 /product="phosphoribosylaminoimidazole-succinocarboxamide
 synthase"
 /protein\_id="CAD02687.1"
 /db\_xref="GI:16503704"
 /db\_xref="GOA:Q8XF69"
 /db\_xref="UniProt/Swiss-Prot:Q8XF69"

/translation="MOKOAEIYRKAKTVYSTENPDLLVEFRNDTSAGDARIEQFD
 RKGVNKNKFNPHINTKLAEGAPIQOMERLLSDTECLYKKLEMPVECVVRNRAAGSLV
 KRLEVEGMEINPFIIDLFLKNDALHDPMNSSYCEFTGWSQENLARMKELTYKAND
 VLKKLFDDAGLILVDFKLEFGIYKGEVVLGDEFSPDSRLMDKETLDKMDKDRFROSL
 GGLIEAYEAVAHRLGVKLD"
 complement(5421..6122)
 /gene="STY2725"
 /note="Pfam match to entry PF01259 SAICAR synt, SAICAR
 synthetase, score 414.20, E-value 1.2e-120"
 complement(5586..5612)
 /gene="STY2725"
 /note="PS01058 SAICAR synthetase signature 2"
 complement(5829..5873)
 /gene="STY2725"

Query Match	61.5%;	Score 461.2;	DB 1;	Length 145050;
Best local Similarity	78.3%;	Pred. No. 5.5e-133;		
Matches	553;	Conservative 0;	Mismatches 153;	Indels 0; Gaps 0;
QY 1	GATCCATAACCCCAACCTATCGAAATATCGAATCTAGAATATATAAACAATTCTTTT	60		
Db 143748	GATCCATAACCAACCAACTATCGAAATCTCGAATTCAGAATTAATAAACATTCTTTT	143807		
QY 61	TTAATGTTCCGTGTCGGGTACTGTCTACCAAAACAGAGAGATTAACAAGTACACCGAC	120		
Db 143808	TTAATATTTCAATCGGGTACTGTCTGTAAACAGAGAGATGACAAAGTACACCGAT	143867		
QY 121	CCTTTAAGTCTTTTGGACTTACACCTGATTAACCGCTATGACCGCAGACCGAACAA	180		
Db 143868	GCTTTAAGCGCTTCTGACTTATACATTAATACACCGCCCTTACCCCGCCCAACAA	143927		
QY 181	TATTCTCGCCCTAGCTCTGTACGTGCGATGATTTCGTAAGTAAACCCGCTGCTGC	240		
Db 143928	TATTCTCGCGCTAGCGCAGCACTGCACATGTTTTTCGACAGATTAATCGTACTGC	143987		
QY 241	AGGATGAGTCTGGGATTTTGTATGTGATCTGTGTGCGGGCATTTTCATTTTCACT	300		
Db 143988	AGTATGAGTCTGGGATTTTGTGTCGTGCTGTGCGCGGGATTCGCTTTCCCT	144047		
QY 301	GGCAGTATGACCCCGCAGCGGTACACCTTTGAGTTGGCGGGCGGCATATATTGT	360		
Db 144048	TGCGGTATGATCGCGCAATCATTCATTGCTCAAGTTGGGTAGGGGCTGCGTATATCCT	144107		
QY 361	CTGCTGCGGTGAAATCGCCACAGCCCAACAAAGAAAGACGGACTTCAGGCCAAACC	420		
Db 144108	GTCGCTGCGGTGAAATTCGCCACAGTCCAGCGCGGATGAGAGGTGACCAAAACC	144167		
QY 421	AATCAGCTTTGGCCGACCTTTGCTTGCAGTTTGTGAACGTCAAAATCATTTGTACGG	480		
Db 144168	GGTTGGCTTTGGGTAAAGCTTTGGCCTGCAGTTTGTGAACGTCAAAATTAATCTGTACGG	144227		
QY 481	TGTTACGCACTGTGACGTTTGTCTGCCGCAACACAGGCGTTAAGCTGGGTAGTTGG	540		
Db 144228	CATTACCGCTGTGTACATTTGTTTACCGCAACGACGCGTGAACCTGGGTACATCGG	144287		
QY 541	CGTCAGCGTTTGTGCGCGCATGATTGGGACGTTTGGCAATGTGTGCGGCGCTGGCGG	600		
Db 144288	CGTCAGTATATTTGCTGCGCATTTGATCGGACGCTTCGGTAATGTGTGCTGGCGCTGGCGG	144347		
QY 601	GCATCTGTTTACGCAATGTTTGCAGATGATGTCGCCAGTTAATATTCGCTTGCCT	660		
Db 144348	ACATCTGTTTACGCGCGGCTTTCGCCATTAATGTCGCCAGTTAATATTAATCTGCGCTT	144407		
QY 661	GTTGCTGCTATTTGCGCGGTACGCAATTTCTATTAACGAAAAAAA 706			
Db 144408	ACTGCTGTGTACTGCGCGGTGCGGATTTTACTGACACAATTA 144453			

RESULT 15  
 AE016835/c 300431 bp DNA linear BCT 21-MAR-2003  
 LOCUS AE016835  
 DEFINITION Salmonella enterica subsp. enterica serovar Typhi Ty2, section 2 of  
 16 of the complete genome.







GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 10, 2005, 07:40:43 ; Search time 178 Seconds  
(without alignments)  
6894.424 Million cell updates/sec

Title: US-10-620-487-1

Perfect score: 750  
Sequence: 1 gatccataaccaccaacta.....ttagtaacttgctacttaag 750

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA:\*

1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PCtUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	265.8	35.4	618	4	US-09-489-039A-5490 Sequence 5490, Ap
2	156	20.8	372	4	US-09-489-039A-5497 Sequence 5497, Ap
3	42	5.6	669	4	US-09-328-352-2345 Sequence 2345, Ap
4	39.8	5.3	636	4	US-09-543-681A-2536 Sequence 2536, Ap
5	39.6	5.3	750	4	US-09-252-991A-9146 Sequence 9146, Ap
6	38.2	5.1	636	4	US-09-489-039A-3271 Sequence 3271, Ap
7	35	4.7	4685	4	US-09-949-016-13848 Sequence 13848, A
8	34.2	4.6	645	4	US-09-489-039A-2622 Sequence 2622, Ap
9	33.4	4.5	202	4	US-09-513-999C-29855 Sequence 29855, A
10	33.4	4.5	1055	3	US-09-410-028-1 Sequence 1, Appli
11	33.4	4.5	3077	3	US-09-949-016-12200 Sequence 12200, A
12	33.4	4.5	27056	4	US-09-949-016-13115 Sequence 13115, A
13	33.4	4.5	27057	4	US-09-949-016-15753 Sequence 15753, A
14	33.2	4.4	236964	4	US-09-949-016-15753 Sequence 15753, A
15	33	4.4	72604	3	US-09-268-992-7 Sequence 7, Appli
16	33	4.4	72604	3	US-09-657-474-7 Sequence 7, Appli
17	33	4.4	194933	4	US-09-949-016-14172 Sequence 14172, A
18	32.8	4.4	1194	4	US-09-252-991A-9124 Sequence 9124, Ap
19	32.8	4.4	1326	4	US-09-252-991A-9206 Sequence 9206, Ap
20	32.8	4.4	1329	4	US-09-252-991A-8853 Sequence 8853, Ap
21	32.8	4.4	234884	4	US-09-949-016-16420 Sequence 16420, A
22	32.6	4.3	505	4	US-09-621-976-15639 Sequence 15639, A
23	32.2	4.3	1001	3	US-09-641-638-340 Sequence 340, App
24	32.2	4.3	1001	4	US-10-170-097-340 Sequence 340, App
25	32.2	4.3	5127	3	US-09-462-606-1 Sequence 1, Appli
26	32.2	4.3	7207	3	US-09-462-606-3 Sequence 3, Appli
27	32.2	4.3	150032	4	US-09-949-016-14321 Sequence 14321, A

C	28	32	4.3	1338	4	US-09-489-039A-6698	Sequence 6698, Ap
C	29	31.8	4.2	789	4	US-09-134-000C-2827	Sequence 2827, Ap
C	30	31.6	4.2	28054	4	US-09-902-540-1248	Sequence 1248, Ap
C	31	31.4	4.2	965	4	US-09-270-767-14185	Sequence 14185, A
C	32	31.4	4.2	197875	4	US-09-949-016-15425	Sequence 15425, A
C	33	31.2	4.2	601	4	US-09-949-016-47364	Sequence 47364, A
C	34	31.2	4.2	601	4	US-09-949-016-64780	Sequence 64780, A
C	35	31.2	4.2	601	4	US-09-949-016-205071	Sequence 205071, A
C	36	31.2	4.2	32010	4	US-09-949-016-13127	Sequence 13127, A
C	37	31.2	4.2	101349	4	US-09-949-016-17433	Sequence 17433, A
C	38	31.2	4.2	218940	4	US-09-949-016-17539	Sequence 17539, A
C	39	31.2	4.2	300402	4	US-09-949-016-13632	Sequence 13632, A
C	40	31.2	4.2	1230025	4	US-09-198-452A-1	Sequence 1, Appli
C	41	31.2	4.2	1230230	4	US-09-438-185A-1	Sequence 1, Appli
C	42	30.8	4.1	726	4	US-09-328-352-3033	Sequence 3033, Ap
C	43	30.8	4.1	936	4	US-09-489-039A-5469	Sequence 5469, Ap
C	44	30.8	4.1	1040	4	US-09-602-787A-305	Sequence 305, App
C	45	30.8	4.1	1497	4	US-09-220-132-94	Sequence 94, Appli

ALIGNMENTS

RESULT 1									
US-09-489-039A-5490									
; Sequence 5490, Application US/09489039A									
; Patent No. 6610836									
; GENERAL INFORMATION:									
; APPLICANT: Gary Breton et. al									
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA									
; FILE REFERENCE: 2709.2004001									
; CURRENT APPLICATION NUMBER: US/09/489, 039A									
; PRIOR FILING DATE: 2000-01-27									
; PRIOR APPLICATION NUMBER: US 60/117,747									
; NUMBER OF SEQ ID NOS: 14342									
; SEQ ID NO 5490									
; LENGTH: 618									
; TYPE: DNA									
; ORGANISM: Klebsiella pneumoniae									
; US-09-489-039A-5490									
Query Match 35.4%; Score 265.8; DB 4; Length 618;									
Best Local Similarity 65.0%; Pred. No. 2.1e-83;									
Matches 393; Conservative 0; Mismatches 212; Indels 0; Gaps 0;									
QY	83	TGCTACCAAAACAGAGAGATACACAGTGCACCGACCTTTTAAGTGTCTTGGACT	142						
Db	1	TGCTGCCAGACAGAGAGAAAACCGTGACCTAACCTGATTAAGCGCTTTTAACT	60						
QY	143	TACACCTGATTACCGCTATGACGCGCAGACCGCAACAATATTCTGCCCTTAGCTCTGCT	202						
Db	61	TACACTTGATGATCAGCGCCCTGACACCCGCGGCCCAACAATATCTGCGCTGACGCGCTC	120						
QY	203	ACGTGCGATGATTGCTCAAGTACCCGCGTGTGCGAGGATGATGAGATTTTG	262						
Db	121	ACCTCCATGCTACGCGCGCAGCTGCGGTGTGCGCGGATGAGCGTGGTTATT	180						
QY	263	ATTGTGATGTTACTGTGTGCGGCAATTTCATTTCACTGCGAGTGAACCCGACGCG	322						
Db	181	ATCACTATGTTGATTGCGCGCTTAACTTTTCTGCTGTTGAACCTGACTCAGCCTTT	240						
QY	323	GTACACCTTTTGAAGTGGCGGCGCATATATTGCTGCTGCGCGTGAATCGCC	382						
Db	241	ACCTGTATGTGATGATGATTTGGCGCGCGTACATCTGTGCTGCGCTGCAAGTTGCC	300						
QY	383	ACGACCCCAACAAGAAGACGACTTCAGGCAAAACCAATCAGCTTTGGCCAGCTTT	442						
Db	301	AAAAGCAACCGGCAAGGACCGCGAGTGTGAACCGGTAGGCTTTGGCGGAGTCTG	360						
QY	443	GCTTGCAGTTTGTGAACGTCAAAATCATTTTGTACGGTGTACGGCACTGTGACGTTT	502						



Db 361 GGGCTACAGTTCGTCAAGCTTAAATATATCTGTATGGCATCACC GCGCTTTCTACCTTT 420  
QY 503 GTTCTGCCGAACACAGCGCTTAAGCTGGGTAAGTGGCGTACGCGTTTGGCGCATG 562  
Db 421 GTTCTGCCGTACTCCGAGCCTGTTGGCTATCAGCGTAGTCTGCTGCGCGGCG 480  
QY 563 ATTGGAGCGTTTGGCAATGTGTGCTGGCGCGCTGGCGGCATCTGTTTCAAGCATTTGTTT 622  
Db 481 ATTGGCGCGCTGGAAATCTGTGCTGGCGCGCTGGCGGCATCTGTTTCAAGCGCTGTTT 540  
QY 623 CGCCAGTATGTCGCCAGTTAAATATCGTCTTGCCCTGTGCTGCTCTTATGCGCGGTA 682  
Db 541 TTGCTGTACGGCGGCACTGAATGATGCTGCTGCGTTATTGCTTACTGTGCTGTG 600  
QY 683 CGCAT 687  
Db 601 CGGAT 605

RESULT 2  
US-09-489-039A-5497  
; Sequence 5497, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489, 039A  
; CURRENT FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 5497  
; LENGTH: 372  
; TYPE: DNA  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-5497

Query Match 20.8%; Score 156; DB 4; Length 372;  
Best Local Similarity 71.8%; Pred. No. 1.4e-44;  
Matches 204; Conservative 0; Mismatches 80; Indels 0; Gaps 0;  
QY 19 TATCGAAATATCGAATCTAGATATATAAAACATTGTTTTTAAATGTTCCGTGCG 78  
Db 85 TATCGATATTATCGAAGCGAAGATGAAAAAATATTCATTTTTTAAATTATACGCTTCG 144  
QY 79 GTACTGTCTACCAAAACAGAGAGATTAACAAGTACACCGACCTTTTAAGTGTCTTTG 138  
Db 145 GTAGTGTCTGCAGACAGAGAGAAAAACCGTGACACCTACCTGATAGAAGCGCTTTT 204  
QY 139 GACTTACACCCGTATACCGCTATGACGCCAGACCGAACAATATCTGCGCCCTTAGCTC 198  
Db 205 AACTTACACTTTGATCACCGCCCTGACACCCCGGCCCAACAATATCTGCGCTGACGAG 264  
QY 199 TGCTACGTGCGATGATTTGTCAAAGTACCCGCGTGTGCGCAGGAGATGATCTGGGATT 258  
Db 265 CGTCACTCCCATGCTCTACGCGCGCAGCCTGCGGCTGCTGGCCGGGATGAGCGTGGGTT 324  
QY 259 TTGATTTGATGTTACTGTGTGCGGGCATTTTCATTTTCACTG 302  
Db 325 TATTATCACTATGTTGATTTGCGCGCTTTAACCTTTTCGCTG 368

RESULT 3  
US-09-328-352-2345  
; Sequence 2345, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC99-03PA

; CURRENT APPLICATION NUMBER: US/09/328,352  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 2345  
; LENGTH: 669  
; TYPE: DNA  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-2345

Query Match 5.6%; Score 42; DB 4; Length 669;  
Best Local Similarity 50.4%; Pred. No. 0.00056;  
Matches 133; Conservative 0; Mismatches 125; Indels 6; Gaps 1;  
QY 105 AACAGTGACACCGACCCCTTTTAAGTCTTTTGGACTTACACCCGTGATTACCGCTATGA 164  
Db 53 AGCATATGGCTTTTAACATTTTATGCAATTTGGAGTGTCTCCATCTTTTATTATTA 112  
QY 165 CGCCAGACCGAACATATTTCTGCCCTTAGCTCTGCTACGTGCAATGGAATTTGTCAAA 224  
Db 113 CGCCAGGGGCGACACTGGGCTACGCCAATTTCCGAGATTAAGGTAAGT-----CG 166  
QY 225 GTACCCCGCTGTGCGCAGGAGATGATCTGGGATTTTGATTTGTATGTTACTGTGCGG 284  
Db 167 TCGTCCCCCTGTGTGCAAGTATGCTATTGCGGCACCTTATTACGATTTATTGTAGCGG 226  
QY 285 GCATTTCAATTTCACTGCGCAGTATGACCCGCGCAGGTACACCTTTTGAGTTGGCGG 344  
Db 227 CTGTTGTGCTGCTGTGTAGCAATAATACCAACTGCTGTGATGATCTTACTGTTGCGG 286  
QY 345 GGGCGCATATATTTGCTGCTGCG 368  
Db 287 GTTCTGCTATTTATTTATGATGG 310

RESULT 4  
US-09-543-681A-2536  
; Sequence 2536, Application US/09543681A  
; Patent No. 6605709  
; GENERAL INFORMATION:  
; APPLICANT: GARY BRETON  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILI  
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.1002-001  
; CURRENT APPLICATION NUMBER: US/09/543,681A  
; CURRENT FILING DATE: 2000-04-05  
; PRIOR APPLICATION NUMBER: US 60/128,706  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 8344  
; SEQ ID NO 2536  
; LENGTH: 636  
; TYPE: DNA  
; ORGANISM: Proteus mirabilis  
US-09-543-681A-2536

Query Match 5.3%; Score 39.8; DB 4; Length 636;  
Best Local Similarity 46.5%; Pred. No. 0.0033;  
Matches 128; Conservative 0; Mismatches 147; Indels 0; Gaps 0;  
QY 111 TGACACCGACCTTTTAAGTCTTTTGGACTTACACCCGTGATTACCGCTATGACGCCAG 170  
Db 32 TGACAAATAGCCTAGTGTGTTGTTGAGCGTTTCCCTTTTATGAGCAATTACGCCCG 91  
QY 171 GACCGAACATATTTCTGCCCTTAGCTCTGCTACGTCGCATGATTTGTCAAAGTACCC 230  
Db 92 GACCTAATATTTACTATTAACTCTTCGGGGCTCATGTTGGTTAAACGCTCTTTAA 151  
QY 231 GCGTGTGCGCAGGATGAGTCTGGGATTTTGAATGTGATGTTACTGTGCGGGCATTT 290  
Db 152 CGTTGATGGCAGGCAATATCGTGGAAATGCAATGTATTTACTTACTGCTGCGTTGGTA 211  
QY 291 CATTTTCACTGGAGTATGACCCGCGCAGCGGTACACCTTTTGAGTTGGCGGGCGG 350  
Db 212 TTGCGACATTTATGATTTATTTATCTGCAATGCAATATCGGTTTAAAGATTTAGTAGCA 271



Qy	351	CAVATAITGTCTGCGTGGCGTGAATAATCGCCAC	385
Db	272	TTTATCTTTGTGTGGTTAGCTTTGAAAAACAGCAACC	306

## RESULT 5

US-09-252-991A-9146  
; Sequence 9146, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 9146  
; LENGTH: 750  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-9146

Query Match	5.3%	Score 39.6;	DB 4;	Length 750;
Best Local Similarity	47.2%;	Pred. No. 0.0043;		
Matches 120; Conservative	0;	Mismatches 134;	Indels 0;	Gaps 0;

OY	131	GCTTTTGGACTTACACCCTGATTAACCGCTATGACGCCAGGACCGAACAAATATTCTCCGC	190
Db	160	GCGTTCTCCCTGTTCGCCCTTGCTCACTTCGGTCAACGCCCGGTCCGAACAACACCATGTGTG	219
OY	191	CTTAGCTCTGCTACGTCGCATGGAATTGCTCAAAGTACCCCGCTGCTGGCAGGATGAGT	250
Db	220	CTGGCCTCGGGGTCMACTTCGGTTTCGTCCGCTCCATCCGCACATCTCGGGATCAGT	279
OY	251	CTGGGATTTTGTATTGTGATGTACTGTGTGCGGGCATTTCATTTTCACTGGCAGTGAAT	310
Db	280	TGCGGCTTCTTCATCATGTGATGCGCGTGGGCTTCGGCCTGGGCAACGGTTCGAGGCC	339
OY	311	GACCCGGCAGCGGTACACCTTTGAAGTTGGCGGGGGCGGCATATATTGTCTGCGTGGCG	370
Db	340	TATCCGGTGCTGTACACCATCTCGCGTACGTCCGGCGGCTAACCCTGCTGTACTTGGCC	399
OY	371	TGGAATAATCGCCAC	384
Db	400	TGGAAGATCGCCAC	413

## RESULT 6

```

US-09-489-039A-3271
: Sequence 3271, Application US/09489039A
: Patent No. 6610836
: GENERAL INFORMATION:
: APPLICANT: Gary Breton et. al
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
: TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: 2709.2004001
: CURRENT APPLICATION NUMBER: US/09/489, 039A
: CURRENT FILING DATE: 2000-01-27
: PRIOR APPLICATION NUMBER: US 60/117,747
: PRIOR FILING DATE: 1999-01-29
: NUMBER OF SEQ ID NOS: 14342
: SEQ ID NO 3271
: LENGTH: 636
: TYPE: DNA
: ORGANISM: Klebsiella pneumoniae
US-09-489-039A-3271

```

Query Match	5.1%;	Score 38.2;	DB 4;	Length 636;
Best Local Similarity	45.9%;	Pred. No. 0.012;		
Matches 130; Conservative	0;	Mismatches 153;	Indels 0;	Gaps 0;

QY	102	GATAACAAGTGACACCGACCCCTTTTAAGTGCCTTTTGGACCTTACACCCCTGAATTAACCGCTA	161
db	29	GATAATTTCATGTCAACAGAAATTCGATCGCATTATTAATTATGCAATTTGTTCACCTCAA	88
QY	162	TCACGCCAGGACCGAACAATAATCTCGCCCTTAGCTCTGCTACGTGCATGGATTTCGTC	221
Db	89	TTACACCCCGGCCTAATAACAATAGTGTGCTTGCCATCAGGTCTGAATTATGTGTTAAAC	148
QY	222	AAAGTACCCGCGTGTGGCAGGATGAGTCTGGATTTTTGATTTGATGTACTGTGTG	281
Db	149	GTAGCCTTCCCATGTCTTGGCATCAGCATTTGTTTGTCTTAATGTCATTCGCCGTTG	208
QY	282	CGGGCATTTCATTTTCATGCGCAGTGAATTGACCCGCGAGCGGTACACCTTTGAGTTGGG	341
Db	209	GTACTGCTCTGTCGTGTGTTGAALATGGCCCTCAGGTGATATAGCCTTACGGATAT	268
QY	342	CGGGGGCGGCATATATTGTCTGCGCTGCGCTGAAAATCGCCAC	384
Db	269	GCGGTGCGACTTATTTAATATATCTGGCATGGAATATAGCCAC	311

## RESULT 7

```

US-09-949-016-13848/C
; Sequence 13848, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13848
; LENGTH: 46885
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(46885)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13848

```

Query Match	4.7%;	Score 35;	DB 4;	Length 46885;
Best Local Similarity	56.1%;	Pred. No. 3.1;		
Matches	87;	Conservative	0;	Mismatches 65;
			Indels 3;	Gaps 1;

OY		3	TCCATAACCCCCTATCGAAATATCGAATCATTAATAAACATTCAATTTT	62
Db		2108	TCTAAAACCAAAGCAGAAGAAAAATACCACAGAANAACAATACTGATTTCT	2049
OY		63	AAATGTTCGT--GTGGGTACTGTCTACCAAAACAGAGAGATPACAAGTGACCGA	119
Db		2048	CCCCTTTCTGTGTGCTTGTATAGGCTACCTGGAGACATCAGAGACAGTTAAAGCCA	1989
OY		120	CCCTTTTAAGTCTTTTGGACTTACACCCTGATT	154
Db		1988	ATCTTTTTTATCAATGAGGTACTTACAGAGATTAATT	1954

## RESULT 8

US-09-489-039A-2622

```
; Sequence 2622, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 2622
; LENGTH: 645
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-2622
```

```
Query Match          4.6%; Score 34.2; DB 4; Length 645;
Best Local Similarity 47.8%; Pred. No. 0.32;
Matches 99; Conservative 0; Mismatches 108; Indels 0; Gaps 0;
```

```
QY 162 TGACGCCAGACCGAACAATATTTCTGCCCTTAGCTCTGCTAGCTCGCATGGATTTCGTC 221
    |||||
Db 80 TGACGCCAGACTCTGTAFACTGCTCTTATTTCCGACGGCGACTGCTGAAGGGGACGCA 139

QY 222 AAAGTACCCGCGTGTGCGAGGAGTGTGCTGGATTGTTGATTTGATGTGACTGTGTG 281
    |||||
Db 140 AAGCCTTGATGCGCGCTTGAAGCATGCATCTCGGCTCTTATCTGGGGCGCGTGTG 199

QY 282 CGGGCATTTCAATTTCACTGGCAGTGTGACCCCGGACCGGTACACCTTTTGAGTTGGG 341
    |||||
Db 200 CTTTGGTTTAAAGTGGCGTGAAGTGGCGGTTTCTGAGTTGGCTTATACCTTATTAATGCT 259

QY 342 CGGGGGCGGCATATATTTGCTGGCTGG 368
    |||||
Db 260 GCGGAGCAGATACCTTTGCTGGCTGG 286
```

```
RESULT 9
US-09-513-999C-29855
; Sequence 29855, Application US/09513999C
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclet, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 29855
; LENGTH: 202
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 173
; OTHER INFORMATION: m=a or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 189
; OTHER INFORMATION: r=a or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 197
; OTHER INFORMATION: m=a or c
US-09-513-999C-29855
```

```
Query Match          4.5%; Score 33.4; DB 4; Length 202;
Best Local Similarity 54.0%; Pred. No. 0.28;
Matches 61; Conservative 3; Mismatches 49; Indels 0; Gaps 0;
```

```
QY 3 TCCATAACCCCAACCTATCGAAATAATCGAATCTAGATAATAAAACATTCAATTTT 62
    |||||
Db 86 TACATAAGTAATAATCTTGATTTCTAAATTAATCCAGATTAACTAATATATATATTATT 145

QY 63 AAATGTCCTGTCGGGTACTGTCTACCAAAACAGAGAGATAACAAGTACA 115
    |||||
Db 146 ATATCTTTGTTGTTATTAATAATGTTTAAACAATAAAATTAARCAATTGAMA 198
```

```
RESULT 10
US-09-410-028-1/c
; Sequence 1, Application US/09410028
; Patent No. 6107471
; GENERAL INFORMATION:
; APPLICANT: ARAHIRA, Masaomi
; APPLICANT: FUKAZAWA, Chikafusa
; TITLE OF INVENTION: PLANT-DERIVED, ASPARAGINE RESIDUE-SPECIFIC ENDOPROTEASE
; FILE REFERENCE: 8361-0007-0
; CURRENT APPLICATION NUMBER: US/09/410,028
; CURRENT FILING DATE: 1999-10-01
; EARLIER APPLICATION NUMBER: JP 327537/1998
; EARLIER FILING DATE: 1998-11-04
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1055
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (85) .. (810)
US-09-410-028-1
```

```
Query Match          4.5%; Score 33.4; DB 3; Length 1055;
Best Local Similarity 55.7%; Pred. No. 0.86;
Matches 64; Conservative 0; Mismatches 51; Indels 0; Gaps 0;
```

```
QY 496 GACGTTGTTCTGCGCCGAACACAGCGCTTAAGCTGGTAGTTGGCGTCAGCGTTTGT 555
    |||||
Db 294 GAGTTGGCATGCGCCCAAGAACAGAGTTGAGGTGAACGATTGGGCAAGATCTGGGT 235

QY 556 GCGGATGATTGGGACGTTTGGCAATGTGTCTGGCGCGCTGGCGGCGCATCTGTT 610
    |||||
Db 234 GAAGATTTTGGAGGTGATTTCACGATGAAGTTGGCGCTGTAGGGGCAATAGCTT 180
```

```
RESULT 11
US-09-410-028-3/c
; Sequence 3, Application US/09410028
; Patent No. 6107471
; GENERAL INFORMATION:
; APPLICANT: ARAHIRA, Masaomi
; APPLICANT: FUKAZAWA, Chikafusa
; TITLE OF INVENTION: PLANT-DERIVED, ASPARAGINE RESIDUE-SPECIFIC ENDOPROTEASE
; FILE REFERENCE: 8361-0007-0
; CURRENT APPLICATION NUMBER: US/09/410,028
; CURRENT FILING DATE: 1999-10-01
; EARLIER APPLICATION NUMBER: JP 327537/1998
; EARLIER FILING DATE: 1998-11-04
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 3077
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
```

```
; NAME/KEY: CDS
; LOCATION: (85)..(321)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (389)..(454)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1160)..(1300)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2414)..(2506)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2707)..(2895)
; US-09-410-028-3
```

```
Query Match      4.5%; Score 33.4; DB 3; Length 3077;
Best Local Similarity 55.7%; Pred. No. 1.8;
Matches 64; Conservative 0; Mismatches 51; Indels 0; Gaps 0;
```

```
QY      496 GACGTTGTTCTGCGCAACACAGCGGTTAGCTGGGTAGTTGGCGTCAGCGTTTGGCT 555
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      294 GAGTTTGCAATGCCCCAAGAACGAGTTTGAGGTGAACGATTGGGCAAGATCTGGGGT 235

QY      556 GCGCATGATGGAGCTTTGGCAATGTGTCTGGGCGCTGGCGGCGCATCTGTTT 610
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      234 GAAGATTTTGGGAGGTGATTCACGATGAGTTGGCGCTGTAGGGGCATAGCTT 180
```

## RESULT 12

```
US-09-949-016-12200
; Sequence 12200, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12200
; LENGTH: 27056
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-12200
```

```
Query Match      4.5%; Score 33.4; DB 4; Length 27056;
Best Local Similarity 58.6%; Pred. No. 7.9;
Matches 58; Conservative 0; Mismatches 41; Indels 0; Gaps 0;
```

```
QY      537 TTGGCGTCAGCGTTTGTGCGCGATGATTGGACGTTTGGCAATGTGTCTGGCGCGCTGG 596
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      3385 TTTGTGTCTGCGTGTCAAGGAGAGCGGGGTGAGGGGTGGCAGTGGGTGCAATGTGGGG 3444

QY      597 CGGGCATCTGTTTTCAGCGATTGTTTGGCCAGTATGCTC 635
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      3445 GGGGGGATATGTCTGGAGCGCTGCCGTCCAGAGGCTC 3483
```

## RESULT 13

```
US-09-949-016-13115
; Sequence 13115, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
```

```
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13115
; LENGTH: 27057
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-13115
```

```
Query Match      4.5%; Score 33.4; DB 4; Length 27057;
Best Local Similarity 58.6%; Pred. No. 7.9;
Matches 58; Conservative 0; Mismatches 41; Indels 0; Gaps 0;
```

```
QY      537 TTGGCGTCAGCGTTTGTGCGCGATGATTGGACGTTTGGCAATGTGTCTGGCGCGCTGG 596
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      3385 TTTGTGTCTGCGTGTCAAGGAGAGCGGGGTGAGGGGTGGCAGTGGGTGCAATGTGGGG 3444

QY      597 CGGGCATCTGTTTTCAGCGATTGTTTGGCCAGTATGCTC 635
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      3445 GGGGGGATATGTCTGGAGCGCTGCCGTCCAGAGGCTC 3483
```

## RESULT 14

```
US-09-949-016-15753
; Sequence 15753, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15753
; LENGTH: 236964
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(236964)
; OTHER INFORMATION: n = A,T,C or G
; US-09-949-016-15753
```

```
Query Match      4.4%; Score 33.2; DB 4; Length 236964;
Best Local Similarity 50.0%; Pred. No. 41;
Matches 83; Conservative 0; Mismatches 83; Indels 0; Gaps 0;
```

```
QY      353 TATATTGTCTGCGTGGCGTGAATAATGCCACAGCCCAACAAGAGACGACTTCAG 412
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      234821 TATAGCGTTTATCAGCTGTGCGCATATTACTGTACATGTTTAATAAGCAAGCAAT 234880

QY      413 GCAAAACCAATCAGCTTTTGGGCCAGCTTTGCTTTCAGCTTGTGCAACGTCAAAATCATT 472
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      234881 CAAAAACCTTTTGTGTTGTAGCCCTGCTTTGCTTCACAATTTGCTTCAATTTGTTT 234940
```

Oy	473	TGTACGCTGTTACGGCACTGTGCAGCTTGTCTTCGCCGAACAC	518
Db	234941	TGTTAATAATAATGCACCCTTGTAATCTTGTACTGTTGCAATAAC	234986

## RESULT 15

```

US-09-268-992-7
; Sequence 7, Application US/09268992
; Patent No. 6342351
; GENERAL INFORMATION:
; APPLICANT: Chen, H.
; APPLICANT: Freimer, N.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CHROMOSOME-18p RELATED DISORDERS
; FILE REFERENCE: 7853-138
; CURRENT APPLICATION NUMBER: US/09/268, 992
; CURRENT FILING DATE: 1999-03-16
; EARLIER APPLICATION NUMBER: 09/236,134
; EARLIER FILING DATE: 1999-01-22
; EARLIER APPLICATION NUMBER: 60/106,056
; EARLIER FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: 60/088,312
; EARLIER FILING DATE: 1998-06-05
; EARLIER APPLICATION NUMBER: 60/078,044
; EARLIER FILING DATE: 1998-03-16
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 72604
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified base
; LOCATION: all n positions
; OTHER INFORMATION: n=a, c, g, or t
US-09-268-992-7

```

Query Match	4.4%	Score 33;	DB 3;	Length 72604;
Best Local Similarity	50.3%;	Pred. No. 21;		
Matches 81; Conservative	0;	Mismatches 80;	Indels 0;	Gaps 0;

QY 47 AACATTCATTTTTTTAAATGTTCCGTCGGGTA CTGCTACCAAAACAGAGGAGATAA 106  
| | | | | | | | | | | | | | | | | | | | | |  
Db 22838 ATACATATATTTCTATATATATACATACACACTTATTTAATAACCAATAGTATCCTTT 22887  
| | | | | | | | | | | | | | | | | | | | | |  
QY 107 CAAGTGACACCGACCTTTTAAAGTCTTTTGGACTTACACCCCTGAATTACCGGTATGACG 166  
| | | | | | | | | | | | | | | | | | | | | |  
Db 22888 CAATGAAAAACAGTAATTTTAACATAAATGAATGAATTAATAATCTAAAGTAAAACTTGACA 22947  
| | | | | | | | | | | | | | | | | | | | | |  
QY 167 CCAGGACCGAACAAATATCTCGCCCTTAGCTCTGCTACGTC 207  
| | | | | | | | | | | | | | | | | | | | | |  
Db 22948 ACAGTGATGCAGAAATTTTTGCTCCCTTAGCTCAGTTAGGTC 22988  
| | | | | | | | | | | | | | | | | | | | | |

Search completed: September 10, 2005, 09:50:53  
Job time : 181 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 10, 2005, 08:56:14 ; Search time 625 Seconds  
(without alignments)  
7880.689 Million cell updates/sec

Title: US-10-620-487-1

Perfect score: 750

Sequence: 1 gatccataaccaccaaaccta.....ttagtaacttgctacttaag 750

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 7351250 seqs, 3283620254 residues

Total number of hits satisfying chosen parameters: 14702500

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:\*

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:\*  
14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:\*  
16: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq:\*  
17: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq:\*  
18: /cgn2\_6/ptodata/1/pubpna/US10F\_PUBCOMB.seq:\*  
19: /cgn2\_6/ptodata/1/pubpna/US10G\_PUBCOMB.seq:\*  
20: /cgn2\_6/ptodata/1/pubpna/US10H\_PUBCOMB.seq:\*  
21: /cgn2\_6/ptodata/1/pubpna/US10I\_PUBCOMB.seq:\*  
22: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*  
23: /cgn2\_6/ptodata/1/pubpna/US11A\_PUBCOMB.seq:\*  
24: /cgn2\_6/ptodata/1/pubpna/US11\_NEW\_PUB.seq:\*  
25: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
26: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	750	100.0	750	US-10-620-487-1	Sequence 1, Appli
2	50.4	6.7	573	US-09-974-300-434	Sequence 434, App
3	39	5.2	247682	US-10-235-192A-28	Sequence 28, Appl
4	36.8	4.9	852	US-10-425-114-35996	Sequence 35996, A
5	36.8	4.9	2727	US-10-425-115-65647	Sequence 65647, A
6	36	4.8	486	US-10-767-795-2906	Sequence 2906, Ap
7	35.2	4.7	474	US-10-259-194A-119	Sequence 119, App

C	8	35.2	4.7	79467	18	US-10-052-482-223	Sequence 223, App
C	9	34.6	4.6	214	9	US-09-933-797-681	Sequence 681, App
C	10	34.6	4.6	3821	10	US-09-814-353-20855	Sequence 20855, A
C	11	34.6	4.6	3825	20	US-10-357-930-25623	Sequence 25623, A
C	12	34.4	4.6	6630	19	US-10-437-963-75212	Sequence 75212, A
C	13	34.2	4.6	2000	17	US-10-260-238-2214	Sequence 2214, Ap
C	14	34.2	4.6	141463	13	US-10-087-192-22	Sequence 22, Appl
C	15	33.8	4.5	497	18	US-10-424-599-84060	Sequence 84060, A
C	16	33.4	4.5	2924	18	US-10-424-599-28594	Sequence 28594, A
C	17	33.4	4.5	9951	17	US-10-257-166-10	Sequence 10, Appl
	18	33.4	4.5	29000	10	US-09-906-158-17	Sequence 17, Appl
	19	33.4	4.5	29000	17	US-10-388-263-466	Sequence 466, App
	20	33.2	4.4	235	18	US-10-424-599-141309	Sequence 141309,
	21	33.2	4.4	826	18	US-10-424-599-97789	Sequence 97789, A
	22	33.2	4.4	3673778	16	US-10-312-841-2	Sequence 2, Appl1
	23	33	4.4	72604	16	US-10-162-497-7	Sequence 7, Appl1
	24	33	4.4	72604	19	US-10-629-313-7	Sequence 7, Appl1
	25	32.8	4.4	64492	18	US-10-378-083-1	Sequence 1, Appl1
C	26	32.8	4.4	414295	20	US-10-719-993-6876	Sequence 6876, Ap
	27	32.6	4.3	451	14	US-10-123-155-126	Sequence 126, App
	28	32.6	4.3	451	15	US-10-146-731-126	Sequence 126, App
	29	32.6	4.3	451	15	US-10-140-472-126	Sequence 126, App
	30	32.6	4.3	451	15	US-10-141-761-126	Sequence 126, App
	31	32.6	4.3	451	16	US-10-142-885-126	Sequence 126, App
	32	32.6	4.3	451	16	US-10-158-790-126	Sequence 126, App
	33	32.6	4.3	451	17	US-10-137-871-126	Sequence 126, App
	34	32.6	4.3	451	17	US-10-140-923-126	Sequence 126, App
	35	32.6	4.3	451	17	US-10-141-756-126	Sequence 126, App
	36	32.6	4.3	451	17	US-10-141-759-126	Sequence 126, App
	37	32.6	4.3	451	17	US-10-140-805-126	Sequence 126, App
	38	32.6	4.3	451	17	US-10-140-864-126	Sequence 126, App
	39	32.6	4.3	451	18	US-10-142-426-126	Sequence 126, App
	40	32.6	4.3	597	20	US-10-363-345A-13041	Sequence 13041, A
C	41	32.6	4.3	597	20	US-10-363-345A-13042	Sequence 13042, A
	42	32.6	4.3	597	21	US-10-363-483A-13041	Sequence 13041, A
C	43	32.6	4.3	597	21	US-10-363-483A-13042	Sequence 13042, A
C	44	32.6	4.3	623	19	US-10-767-701-883	Sequence 883, App
	45	32.6	4.3	858	20	US-10-363-345A-21325	Sequence 21325, A

ALIGNMENTS

RESULT 1  
US-10-620-487-1  
; Sequence 1, Application US/10620487  
; Publication No. US20040038352A1  
; GENERAL INFORMATION:  
; APPLICANT: MAIER, THOMAS  
; TITLE OF INVENTION: METHOD FOR FERMENTATIVE PRODUCTION OF AMINO ACIDS AND AMINO ACID  
; TITLE OF INVENTION: DERIVATIVES OF THE PHOSPHOGLYCERATE FAMILY  
; FILE REFERENCE: MAIER, T-2  
; CURRENT APPLICATION NUMBER: US/10/620,487  
; CURRENT FILING DATE: 2003-07-16  
; PRIOR APPLICATION NUMBER: GERMAN NO. 102 32 930.3  
; PRIOR FILING DATE: 2002-07-19  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1  
; LENGTH: 750  
; TYPE: DNA  
; ORGANISM: Escherichia coli  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (110)..(694)  
; OTHER INFORMATION:  
US-10-620-487-1

Query Match 100.0%; Score 750; DB 18; Length 750;  
Best Local Similarity 100.0%; Pred. No. 8.3e-239;  
Matches 750; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GATCCATAACCCAAACCTATCGAATAATATCGAATCTAGAAATATAAAACATTCAATTTT 60

Db	1	GATCCATAACCCCAACCTATCGAAATATCGAATCTAGATATATAAACATTCATTTT	60
QY	61	TTAAATGTTCCGTGCGGTACTGTCTACCAAAACAGAGAGATACAACTGACACCGAC	120
Db	61	TTAAATGTTCCGTGCGGTACTGTCTACCAAAACAGAGAGATACAACTGACACCGAC	120
QY	121	CCTTTTAAGTGTCTTTTGGACCTTACACCCCTGATTACCGCTATGACGCCAGAACAA	180
Db	121	CCTTTTAAGTGTCTTTTGGACCTTACACCCCTGATTACCGCTATGACGCCAGAACAA	180
QY	181	TATTCCTGCCCTTAGCTCTGTCACTGCGATGAGATTCGTCAAAGTACCCGCTGCTGC	240
Db	181	TATTCCTGCCCTTAGCTCTGTCACTGCGATGAGATTCGTCAAAGTACCCGCTGCTGC	240
QY	241	AGGATGAGTCTGGGATTTTGTATGTGATGTACTGTGTGCGGCGCATTTTCACT	300
Db	241	AGGATGAGTCTGGGATTTTGTATGTGATGTACTGTGTGCGGCGCATTTTCACT	300
QY	301	GGCAGTGATTGACCCGCGCAGCGGTACACCTTTTGAGTTGGCGGGCGGCATATATTGT	360
Db	301	GGCAGTGATTGACCCGCGCAGCGGTACACCTTTTGAGTTGGCGGGCGGCATATATTGT	360
QY	361	CTGGCTGGCGTGAAAAATCGCCACCAAGCCCAACAAAGGAAGACCGACTTCAGGCCAAACC	420
Db	361	CTGGCTGGCGTGAAAAATCGCCACCAAGCCCAACAAAGGAAGACCGACTTCAGGCCAAACC	420
QY	421	AATCAGCTTTTGGGCCAGCTTTGCTTGACGTTTGTGAACGTCAAATCATTTTGTACCG	480
Db	421	AATCAGCTTTTGGGCCAGCTTTGCTTGACGTTTGTGAACGTCAAATCATTTTGTACCG	480
QY	481	TGTTACGGCACTGTGCACTTTGTTCTGCGCAAAACACAGCGCTTAAAGCTGGTAGTTGG	540
Db	481	TGTTACGGCACTGTGCACTTTGTTCTGCGCAAAACACAGCGCTTAAAGCTGGTAGTTGG	540
QY	541	CGTCAGCGTTTGTGCTGCGATGATTGGGACGTTTGGCAATGTGTGCTGGCGCTGGCGGG	600
Db	541	CGTCAGCGTTTGTGCTGCGATGATTGGGACGTTTGGCAATGTGTGCTGGCGCTGGCGGG	600
QY	601	GCATCTGTTTCAGCGATGTTTGGCCAGTATGTCGCCAGTTAATATCGTCTGCCCT	660
Db	601	GCATCTGTTTCAGCGATGTTTGGCCAGTATGTCGCCAGTTAATATCGTCTGCCCT	660
QY	661	GTTGCTGCTATTTGCGCGGTACGCAATTTCTATTAAAGAAAAAAGCGGAAGAGTCCG	720
Db	661	GTTGCTGCTATTTGCGCGGTACGCAATTTCTATTAAAGAAAAAAGCGGAAGAGTCCG	720
QY	721	CCTCTTCGCGCTAGTAACTTGCTACTTAAG	750
Db	721	CCTCTTCGCGCTAGTAACTTGCTACTTAAG	750

RESULT 2

US-09-974-300-434

Sequence 434, Application US/09974300

Patent No. US20020146721A1

GENERAL INFORMATION:

APPLICANT: Berka, Randy M.

APPLICANT: Clausen, Ib Groth

TITLE OF INVENTION: Methods For Monitoring Multiple Gene

TITLE OF INVENTION: Expression

FILE REFERENCE: 10085.500-US

CURRENT APPLICATION NUMBER: US/09/974,300

CURRENT FILING DATE: 2001-10-05

PRIOR APPLICATION NUMBER: 09/680,598

PRIOR FILING DATE: 2000-10-06

PRIOR APPLICATION NUMBER: 60/279,526

PRIOR FILING DATE: 2001-03-27

NUMBER OF SEQ ID NOS: 8481

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 434

LENGTH: 573

TYPE: DNA

```
; ORGANISM: Bacillus licheniformis
US-09-974-300-434

Query Match      6.7%; Score 50.4; DB 9; Length 573;
Best Local Similarity 46.6%; Pred. No. 1.5e-05;
Matches 194; Conservative 0; Mismatches 221; Indels 1; Gaps 1;

OY      124 TTTAAGTCCTTTTGGACTTAACACCCTGATTAACCGCTATGACGCCAGCAACAATAT 183
        ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB       6  TATTATAGCCTTTCTTTCCGTATGTGATTATGACATCCATTACGCCCGCCGAGCAATAT 65

OY      184 TCTCGCCCTTAGCTCTGCTACGTCGCATGATTTGCTCAAAGTACCCCGCTGGCAGG 243
        ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB       66  TTTAATGATGAATGAAGCCGAAAGGTTCCGGCTTTACAGGTTCAATGCGCTTTAGCACGG 125

OY      244 GATGAGTCTGGGATTTTGATGTGATGTTACTGTGTGCGGGCATTTCAATTTCACTGGC 303
        ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB       126 TATCTTGCGCGGGTTTGACGACTTGGGATTTCAAGCGGTCTTTAACATCAGCCTGTA 185

OY      304 AGTGATGACCCGCGCAGCGGTACACCTTTTGAGTTGGCGGGCGGCATATATGTCG 363
        ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB       186 CAATTGATTCGCCGTGTGAGAGCCTTATTTTAACTTGCCGCGCGTGTATTGATTTA 245

OY      364 GCTGGCGTGGAAAATCGCCACCAAGCCCAACAAGAAGACGACTTCAGCAAAAACCAAT 423
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB       246 TTGGCTTTGCAAGTCGGCTTTACCAAAAATAAAAAACAGATTCACAGAAGCCCGCTC 305

OY      424 CAGCTTTGGGGCCAGCTTTCCTTTGCAGTTTGTAACGTCAAAATCATTTTGTACGGTGT 483
        ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB       306 CTCTTTATATCCGGCTTTATATTTTCAAGCTGATCAATATTAAGAAGCATTTTGTCTTCAT 365

OY      484 T-ACGGCACTGTCGACGTTTGTCTGCGCGCAAAACAGAGCGTTAAGCTGGGTAGTT 538
        ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB       366 TAACCGTAATGAGCGCGTTCGTTTGGCGGTTCACCACTTCCTTGAAATCGACAATT 421


RESULT 3
US-10-235-192A-28
; Sequence 28, Application US/10235192A
; Publication No. US20040043389A1
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Jeanette
; TITLE OF INVENTION: Methods and Compositions for Identifying
; TITLE OF INVENTION: Risk Factors for Abnormal Lipid Levels and the Diseases
; TITLE OF INVENTION: and Disorders Associated Therewith
; FILE REFERENCE: MMJ-011
; CURRENT APPLICATION NUMBER: US/10/235,192A
; CURRENT FILING DATE: 2002-09-04
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 247682
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 145329-145428, 187896-187995, 195894-195993
; OTHER INFORMATION: N = any nucleotide
US-10-235-192A-28

Query Match      5.2%; Score 39; DB 18; Length 247682;
Best Local Similarity 55.6%; Pred. No. 2.9;
Matches 75; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

OY      6  ATAACCCCAAACTATCGAAAATATCGAATCTAGAAATATAAAAAACATTCAATTTTTTAA 65
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB     168128 ATAAACCTATATCTATGAATAATAATTGAATTAATAATTAAACCTTCTCACAAAGAA 168187

OY      66 TGTTCCGTGTGGGTACTGTCTACCAAAACAGAGAGATTAACAAGTACACCGACCTTT 125
        ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB     168188 ATGCCAGTCCCAAGATGTGAATTTATCAAAAATGTAAAGAAGAAATTAACAATAATCCTAT 168247

OY      126 TAAGTCTTTTGA 140
```





```
; APPLICANT: lange, Markus B.
; APPLICANT: Ghassemian, Majid
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiyaki
; APPLICANT: Kreps, Joel
; APPLICANT: Moughamer, Todd
; APPLICANT: Provart, Nicholas
; APPLICANT: Rieke, Darrell
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: IDENTIFICATION AND CHARACTERIZATION OF PLANT GENES
; FILE REFERENCE: 70029-NP
; CURRENT APPLICATION NUMBER: US/10/259,194A
; CURRENT FILING DATE: 2003-01-07
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/370,743
; PRIOR FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 662
; SOFTWARE: PatentList.pl version 3.0.4 (C) 2001 Syngenta
; SEQ ID NO 119
; LENGTH: 474
; TYPE: DNA
; ORGANISM: Oryza sativa
; US-10-259-194A-119
```

```
Query Match      4.7%; Score 35.2; DB 17; Length 474;
Best Local Similarity 49.5%; Pred. No. 1.6;
Matches 91; Conservative 0; Mismatches 93; Indels 0; Gaps 0;
```

```
QY 486 CGGCACGTGTCGACGTTTGTCTGCGCAACACAGCGCTTAAGCTGGGTAGTTGGCGTCA 545
    |||||
Db 81 CGGCGCTGCCGCTGCTGTGTCTCCACCGTCACCGCCAGCATGACGAACTTTGGCGACA 140
QY 546 GCGTTTGTCTGGCGATGTGGACGTTTGGCAATGTGTGCGCTGGCGCGGCATC 605
    |||||
Db 141 ATCCGCTGCTGCTGTGTGTGATGATGTGATGATGATGATGATGATGATGATGATG 200
QY 606 TGTTCAGCGATGTTTGGCCAGTATGTGCGCAGTTAATATCGTGTGCCCTGTGTC 665
    |||||
Db 201 TGATTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 260
QY 666 TGGT 669
    |||
Db 261 TGAT 264
```

```
RESULT 8
US-10-052-482-223/c
```

```
; Sequence 223, Application US/10052482
; Publication No. US20040072264A1
; GENERAL INFORMATION:
; APPLICANT: Engelhard, Eric
; APPLICANT: Morris, David
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: A-71087/RMS/DCF
; CURRENT APPLICATION NUMBER: US/10/052,482
; CURRENT FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 241
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 223
; LENGTH: 79467
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
```

```
; NAME/KEY: misc feature
; LOCATION: (4095)..(4369)
; OTHER INFORMATION: "n" at positions 4099 to 4369 can be any base
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (5502)..(5521)
; OTHER INFORMATION: "n" at positions 5502 to 5521 can be any base
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (10089)..(10620)
; OTHER INFORMATION: "n" at positions 10089 to 10620 can be any base
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (13273)..(13370)
; OTHER INFORMATION: "n" at positions 13273 to 13370 can be any base
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (20762)..(20781)
; OTHER INFORMATION: "n" at positions 20762 to 20781 can be any base
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (30751)..(30916)
; OTHER INFORMATION: "n" at positions 30751 to 30916 can be any base
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (46579)..(46772)
; OTHER INFORMATION: "n" at positions 46579 to 46772 can be any base
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (77382)..(77401)
; OTHER INFORMATION: "n" at positions 77372 to 77401 can be any base
; US-10-052-482-223
```

```
Query Match      4.7%; Score 35.2; DB 18; Length 79467;
Best Local Similarity 51.2%; Pred. No. 28;
Matches 82; Conservative 0; Mismatches 78; Indels 0; Gaps 0;
```

```
QY 428 TTTGGGCGCAGCTTGTCTTGTGACGTTTGAAGTCAAATCATTTGTACGGTGTACG 487
    |||||
Db 60135 TTTATGCCCGATGAAGAACTAGCATGTGTGCGCAGGCGTGTATGCGTACAGAGCTGG 60076
QY 488 GCACGTGCGAGTTTGTCTGTGCGCCGAACACAGCGGCTTAAGCTGGGTAGTTGGCGTACG 547
    |||||
Db 60075 GAGCAGACACTGTCCGTTTCTGTGCGCCAGCCAGACCATGGCCAGGGAATGAAGAAAGC 60016
QY 548 GTTTGTGCGCATGATTGGGACGTTTGGCAATGTGTGCT 587
    |||||
Db 60015 ATTTCTACAGCTGTTGTGACTTCAAGGCATGTGTGCT 59976
```

```
RESULT 9
US-09-933-797-681/c
```

```
; Sequence 681, Application US/09933797
; Patent No. US20020155119A1
; GENERAL INFORMATION:
; APPLICANT: Robert A. Sikes et al.
; TITLE OF INVENTION: Isolation and Use of Fetal Urogenital
; FILE REFERENCE: 9901-007-999
; CURRENT APPLICATION NUMBER: US/09/933,797
; CURRENT FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: US/09/482,933
; PRIOR FILING DATE: 2000-01-14
; PRIOR APPLICATION NUMBER: PCT/US99/10746
; PRIOR FILING DATE: 1999-05/14
; PRIOR APPLICATION NUMBER: 60/085,383
; PRIOR FILING DATE: 1998-05-14
; NUMBER OF SEQ ID NOS: 811
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 681
; LENGTH: 214
; TYPE: DNA
; ORGANISM: Murine
; FEATURE:
```



Query Match 4.6%; Score 34.4; DB 19; Length 6630;  
Best Local Similarity 54.8%; Pred. No. 13;  
Matches 68; Conservative 0; Mismatches 56; Indels 0; Gaps 0;  
QY 528 GCTGGTAGTTGGCGTCACCGTTTGTCTGGCGCATGATTGGACGTTTGGCAATGTGTGCT 587  
Db 3310 GCGGTGGCGTTGGCGTCACGAGTGGCGCGGTGGAGTTGGCGCGCGCGGGGTGCG 3251  
QY 588 GGGCGCTGGCGGGCATCTGTTTTCAGCGATTGTTCCAGATATGTCCGCAATTAATA 647  
Db 3250 GTCCGAGCGCGCGCGGAGGTGGCAGCGGTGTGGCGCGCGCATGGCGATGGCGCGAAGTC 3191  
QY 648 TCGT 651  
Db 3190 CCGT 3187

RESULT 13  
US-10-260-238-2214/c  
; Sequence 2214, Application US/10260238  
; Publication No. US20040016025A1  
; GENERAL INFORMATION:  
; APPLICANT: Budworth, Paul R.  
; APPLICANT: Moughamer, Todd G.  
; APPLICANT: Briggs, Steven P.  
; APPLICANT: Cooper, Bret  
; APPLICANT: Glazebrook, Jane  
; APPLICANT: Goff, Stephen A.  
; APPLICANT: Katagiri, Fumiyaki  
; APPLICANT: Kreps, Joel  
; APPLICANT: Provart, Nicholas  
; APPLICANT: Ricke, Darrell  
; APPLICANT: Zhu, Tong  
; TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION  
; FILE REFERENCE: 60111-NP  
; CURRENT APPLICATION NUMBER: US/10/260,238  
; CURRENT FILING DATE: 2002-09-26  
; PRIOR APPLICATION NUMBER: US 60/325,448  
; PRIOR FILING DATE: 2001-09-26  
; PRIOR APPLICATION NUMBER: US 60/325,277  
; PRIOR FILING DATE: 2001-09-26  
; PRIOR APPLICATION NUMBER: US 60/370,620  
; PRIOR FILING DATE: 2002-04-04  
; NUMBER OF SEQ ID NOS: 6077  
; SEQ ID NO 2214  
; LENGTH: 2000  
; TYPE: DNA  
; ORGANISM: Oryza sativa  
; FEATURE:  
; NAME/KEY: N region  
; LOCATION: (1919)..(1919)  
; OTHER INFORMATION: n = any nucleotide  
US-10-260-238-2214

Query Match 4.6%; Score 34.2; DB 17; Length 2000;  
Best Local Similarity 56.8%; Pred. No. 7.6;  
Matches 63; Conservative 0; Mismatches 48; Indels 0; Gaps 0;  
QY 543 TCAGCGTTTGTCTGGCGATGATTGGACGTTTGGCAATGTGTCTGGCGCTGGCGGGC 602  
Db 1871 TGAGCGAATTGGCTGGTTTGGGACTTTTGGGGTGGCTGTGAGCTGTGTATTGGC 1812  
QY 603 ATCTGTTTACGCGATTGTTTGGCCAGTATGTGCGCAGTTAATATCGTGC 653  
Db 1811 GTCCCATCTGGCAGTGTGGGAGGTTGGGTGTGGCCCAATGTCGTGC 1761

RESULT 14  
US-10-087-192-22/c  
; Sequence 22, Application US/10087192  
; Publication No. US20020182586A1  
; GENERAL INFORMATION:  
; APPLICANT: Morris, David W.

; APPLICANT: Engelhard, Eric K.  
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: CANCER  
; FILE REFERENCE: 529452000122  
; CURRENT APPLICATION NUMBER: US/10/087,192  
; CURRENT FILING DATE: 2002-03-01  
; PRIOR APPLICATION NUMBER: US 09/747,377  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: US 09/798,586  
; PRIOR FILING DATE: 2001-03-02  
; NUMBER OF SEQ ID NOS: 2059  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 22  
; LENGTH: 141463  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-087-192-22

Query Match 4.6%; Score 34.2; DB 13; Length 141463;  
Best Local Similarity 47.1%; Pred. No. 83;  
Matches 105; Conservative 0; Mismatches 118; Indels 0; Gaps 0;  
QY 6 ATAAACCCAAACCTATCGAAAATATCGAATCTAGATATATAAAACATTCAATTTTAA 65  
Db 25440 ATCAGCCTCAAAACACTAACCCACAAAGAAAGTTGAATGTAAACAGTACATATTTTAA 25381  
QY 66 TGTCCGTGTGGGTACTGTCTACCAAAACAGAGAGATAACAGTACACCGACCCCTT 125  
Db 25380 TATTAACCTTTAAACATGTTTAAAGTTAAATTGCAGTGCAGTAACTTACCTAA 25321  
QY 126 TAAGTCTTTTGGACTTACACCCGTGATTAACCGTATGACGCGCAGACCGAACAATATTC 185  
Db 25320 AAGTGCAATGCTTGTGACTGCAATTCGCAATTTGTTCCAAAGAAACAATTCAAGATC 25261  
QY 186 TCGCCCTTAGCTCTGCTACGTCGATGATTTGCTCAAGTAC 228  
Db 25260 TGTGCAAACTTTAGTTAGAATATATGCACTTCATCACAGTAC 25218

RESULT 15  
US-10-424-599-84060/c  
; Sequence 84060, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 84060  
; LENGTH: 497  
; TYPE: DNA  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_46923C.1  
US-10-424-599-84060

Query Match 4.5%; Score 33.8; DB 18; Length 497;  
Best Local Similarity 55.6%; Pred. No. 4.7;  
Matches 65; Conservative 0; Mismatches 52; Indels 0; Gaps 0;  
QY 257 TTTTGTATGTGATGTTACTGTGTGCGGCATTTCAATTTCACTGGCAGTGTGACCCG 316  
Db 341 TGTATGCTGTGCTGTGGCTGTGTCTGTAAGCCGCAACACTGGCAGAAATGTCTCG 282  
QY 317 GCAAGCGTACACCTTTTGAAGTGGCGGGGCGGCATATATTTGTCTGCTGGCGTGG 373  
Db 281 GGAAGCTGTCCATTTTCTTGTCTGCAAGGTGGCGGATTTGTATTCTAGGATTGCAAGTGG 225



Search completed: September 10, 2005, 11:37:07  
Job time : 629 secs

---

***This Page Blank (uspto)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using bw model

Run on: September 10, 2005, 07:37:28 ; Search time 3088 Seconds  
(without alignments)  
9244.884 Million cell updates/sec

Title: US-10-620-487-1

Perfect score: 750

Sequence: 1 gatccataaccaccaacta.....ttagtaacttgctacttaag 750

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_hic:\*  
4: gb\_est3:\*  
5: gb\_est4:\*  
6: gb\_est5:\*  
7: gb\_est6:\*  
8: gb\_gss1:\*  
9: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	473	63.1	681	6	CA366654 642278 NC CA366654 642278 NC
2	351	46.8	856	9	CL650636 PRI0110a CL650636 PRI0110a
3	331.4	44.2	791	9	CL688782 PRI014b CL688782 PRI014b
4	321.6	42.9	736	9	CL688905 PRI014d CL688905 PRI014d
5	113.2	15.1	118	6	CA374430 648800 NC CA374430 648800 NC
6	55.2	7.4	819	8	BH447440 BOHQB49TF BH447440 BOHQB49TF
7	50	6.7	420	8	BH374397 AG-ND-180 BH374397 AG-ND-180
8	49.4	6.6	862	9	CL666889 PRI0153C CL666889 PRI0153C
9	43.4	5.8	450	8	BH379453 AG-ND-163 BH379453 AG-ND-163
c 10	39.2	5.2	938	9	CL139330 ISB1-112E CL139330 ISB1-112E
c 11	38.2	5.1	474	9	AG240288 Lotus cor AG240288 Lotus cor
12	38.2	5.1	1101	9	CNS00EBC AL069085 Drosophil AL069085 Drosophil
c 13	37.6	5.0	553	5	BM965274 kj93h10.y BM965274 kj93h10.y
c 14	37.6	5.0	568	5	BQ095448 k005d05.y BQ095448 k005d05.y
c 15	37.6	5.0	583	5	BM965121 kj92c03.y BM965121 kj92c03.y
c 16	37.6	5.0	623	5	BM965443 kj95h06.y BM965443 kj95h06.y
c 17	37.6	5.0	637	5	BM965333 kj94f01.y BM965333 kj94f01.y
c 18	37.4	5.0	493	9	CE114952 tigr-gss- CE114952 tigr-gss-
19	37.4	5.0	884	9	CNS006U0 AL065923 Drosophil AL065923 Drosophil
c 20	37.2	5.0	939	9	CNS00CNG AL059400 Drosophil AL059400 Drosophil
21	36.8	4.9	614	5	BQ628514 sap47h10. BQ628514 sap47h10.
22	36.8	4.9	650	6	CD598566 RK114A3B1 CD598566 RK114A3B1
23	36.8	4.9	650	6	CD599088 RK114A4B0 CD599088 RK114A4B0
c 24	36.8	4.9	712	9	BX163185 Danio rer BX163185 Danio rer

25	36.8	4.9	778	9	BX175544	BX175544 Danio rer
26	36.6	4.9	408	8	BZ716786	BZ716786 OGEBE26TM
27	36.6	4.9	589	9	CC604749	CC604749 OGUHO32TH
c 28	36.6	4.9	787	9	CNS016H3	AL106737 Drosophil
c 29	36.6	4.9	935	9	CC604756	CC604756 OGUHO32TV
30	36.6	4.9	963	9	CG050339	CG050339 PU1IH51TB
c 31	36.4	4.9	440	6	CA678514	CA678514 wlm12.pk0
c 32	36.4	4.9	566	6	CD347163	CD347163 EteStef11
c 33	36.4	4.9	619	5	BM964489	BM964489 kj84e08.y
c 34	36.2	4.8	825	9	CC534336	CC534336 CH240_412
c 35	36	4.8	630	5	BM964821	BM964821 kj88g02.y
c 36	36	4.8	688	8	BZ659718	BZ659718 OGCCC31TM
c 37	35.8	4.8	1181	9	AG127050	AG127050 Pan trogl
c 38	35.6	4.7	492	1	AI908572	AI908572 RC-BT177-
c 39	35.4	4.7	699	9	CE799079	CE799079 tigr-gss-
40	35.4	4.7	733	8	BZ389118	BZ389118 EINDJ32TF
c 41	35.4	4.7	792	9	CG872981	CG872981 ZMIBC028
c 42	35.2	4.7	226	4	BG370992	BG370992 dacl1a02.
43	35.2	4.7	500	5	BP188868	BP188868 BP188868
44	35.2	4.7	500	5	BP189161	BP189161 BP189161
c 45	35.2	4.7	552	9	BX163834	BX163834 Danio rer

ALIGNMENTS

RESULT 1  
CA366654/c 681 bp mRNA linear EST 06-NOV-2002  
LOCUS 642278 NCCCW 1RT Oncorhynchus mykiss cDNA clone 1RT18B19\_B\_A10 5',  
DEFINITION mRNA sequence.

ACCESSION CA366654 GI:24677298  
VERSION CA366654.1  
KEYWORDS EST.  
SOURCE Oncorhynchus mykiss (rainbow trout)  
ORGANISM Oncorhynchus mykiss

REFERENCE  
AUTHORS Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei;  
Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.  
1 (bases 1 to 681)  
Rexroad, C.E. 3rd, Lee, Y., Keele, J.W., Karamycheva, S., Brown, G.,  
Koop, B., Gahr, S.A., Palti, Y. and Quackenbush, J.

TITLE  
Sequence analysis of a rainbow trout cDNA library and creation of a  
gene index

JOURNAL  
COMMENT Cytoenet. Genome Res. 102 (1-4), 347-354 (2003)  
Contact: Rexroad CE  
USDA, ARS, National Center for Cool and Cold Water Aquaculture  
11876 Leetown Road, Kearneysville, WV 25430, USA  
Tel: 304 724 8340 x2129  
Fax: 304 725 0351

FEATURES  
source  
1. 681  
/organism="Oncorhynchus mykiss"  
/mol\_type="mRNA"  
/db\_xref="taxon:8022"  
/clone="1RT18B19\_B\_A10"  
/issue\_type="pooled"  
/lab\_host="DH10B"  
/clone\_lib="NCCCW 1RT"  
/note="Vector: PCMV SPOR6; Site 1: NotI; Site 2: SalI;  
Library made from pooled tissue from brain, gill, liver,  
spleen, muscle, and kidney."

ORIGIN

Query Match 63.1%; Score 473; DB 6; Length 681;  
Best Local Similarity 100.0%; Pred. No. 1.8e-132;  
Matches 473; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



QY	1	GATCCATAACCCCAAAACCTATCGAAAATATCGAAATCTAGAATATAAAAACATTCAATTTT	60
Db	473	GATCCATAACCCCAAAACCTATCGAAAATATCGAAATCTAGAATATAAAAACATTCAATTTT	414
QY	61	TTAAATGTCCGTGTGGGTACTGTCTACCAAAAAGAGAGATTAACAAGTGACACCGAC	120
Db	413	TTAAATGTCCGTGTGGGTACTGTCTACCAAAAAGAGAGATTAACAAGTGACACCGAC	354
QY	121	CCTTTTAAGTGTCTTTTGACTTACACCCGTATTAACCGCTATGACGCCAGACCGAACAA	180
Db	353	CCTTTTAAGTGTCTTTTGACTTACACCCGTATTAACCGCTATGACGCCAGACCGAACAA	294
QY	181	TATTTCTGCCCTTAGCTCTGCTACAGTCGATGGAATTCGTAAGTACCCGCGTGGC	240
Db	293	TATTTCTGCCCTTAGCTCTGCTACAGTCGATGGAATTCGTAAGTACCCGCGTGGC	234
QY	241	AGGATGAGTCTGGGATTTTGTGATTGTGATGTTACTGTGTGCGGGCATTTCAATTTCACT	300
Db	233	AGGATGAGTCTGGGATTTTGTGATTGTGATGTTACTGTGTGCGGGCATTTCAATTTCACT	174
QY	301	GGCAGTGAATGACCCGGGACACCGGTACACCTTTTGAGTTGGCGGGGGCGGCATATATTTGT	360
Db	173	GGCAGTGAATGACCCGGGACACCGGTACACCTTTTGAGTTGGCGGGGGCGGCATATATTTGT	114
QY	361	CTGGCTGGCGTGAAAAATCGCCACACGACCCCAACAAAGAGAAGCGGACTTCAGGCAAAAC	420
Db	113	CTGGCTGGCGTGAAAAATCGCCACACGACCCCAACAAAGAGAAGCGGACTTCAGGCAAAAC	54
QY	421	AATCAGCTTTTGGGCCAGCTTTGCTTTGACGTTGTGAACGTCAAAATCATTT	473
Db	53	AATCAGCTTTTGGGCCAGCTTTGCTTTGACGTTGTGAACGTCAAAATCATTT	1
RESULT 2			
LOCUS	CL650636	856 bp	DNA linear
DEFINITION	PR10110a_C10 - PR10110a.B21 (856) Mixed stage fosmid library of P. pacificus var. California Pristionchus pacificus genomic, genomic survey sequence.		GSS 09-JUL-2004
ACCESSION	CL650636		
VERSION	CL650636.1	GI:50129186	
KEYWORDS	GSS.		
SOURCE	Pristionchus pacificus		
ORGANISM	Pristionchus pacificus		
REFERENCE	Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida; Neodiplogasteridae; Pristionchus.		
AUTHORS	1 (bases 1 to 856)		
TITLE	Srinivasan,J., Otto,G.W., Kahlow,U., Geisler,R. and Sommer,R.J.		
JOURNAL	ApbADB: an Acedb database for the nematode satellite organism		
COMMENT	Pristionchus pacificus		
	Nucleic Acids Res. 32 (1), D421-D422 (2004)		
	Contact: Sommer RJ		
	Evolutionary Biology		
	Max-Planck-Institute for Developmental Biology		
	Spemannstr. 37-39, Tuebingen D-72076, Germany		
	Tel: 00497071601371		
	Fax: 00497071601498		
	Email: ralf.sommer@tuebingen.mpg.de		
	This library was generated at Caltech, Pasadena, USA and end		
	sequenced at Vancouver, Canada.		
	Seq primer: T7		
	Class: fosmid ends.		
FEATURES			
source	Location/Qualifiers		
	1..856		
	/organism="Pristionchus pacificus"		
	/mol_type="genomic DNA"		
	/strain="California"		
	/db_xref="taxon:54126"		
	/clone_lib="Mixed stage fosmid library of P. pacificus		
	var. California"		
	/note="Vector: pDp1fos-5 Fosmid vector"		

Query Match	46.8%;	Score 351;	DB 9;	Length 856;
Best Local Similarity	100.0%;	Pred. No. 2.7e-95;		
Matches 351;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	400	AGACGGACTTCAGGCAAAACCAATCAGCTTTGGGCCACGCTTGTCTTGTGCAGTTTGTGAA	459	
Db	1	AGACGGACTTCAGGCAAAACCAATCAGCTTTGGGCCACGCTTGTCTTGTGCAGTTTGTGAA	60	
QY	460	CGTCAAAATCATTTTGTATACGGGTGTACGGCACTGTGACGTTTGTCTGCGGCAAAACACA	519	
Db	61	CGTCAAAATCATTTTGTATACGGGTGTACGGCACTGTGACGTTTGTCTGCGGCAAAACACA	120	
QY	520	GGCGTTAAGCTGGGTAGTTGGCGCTCAGCGTTTGTCTGGCGATGATTGGGACGTTTGGCAA	579	
Db	121	GGCGTTAAGCTGGGTAGTTGGCGCTCAGCGTTTGTCTGGCGATGATTGGGACGTTTGGCAA	180	
QY	580	TGTGTCTGGGCGCTGGCGGGGCATCTGTTTCAGCGATTGTTTCGCCAGTATGGTGCCA	639	
Db	181	TGTGTCTGGGCGCTGGCGGGGCATCTGTTTCAGCGATTGTTTCGCCAGTATGGTGCCA	240	
QY	640	GTTAATATATCGTCTTGCCCTGTGCTGTGCTATTTGCCGGGTACGCATTTTCTAATTAAAG	699	
Db	241	GTTAATATATCGTCTTGCCCTGTGCTGTGCTATTTGCCGGGTACGCATTTTCTAATTAAAG	300	
QY	700	AAAAAAGCGGAGAGAGTCCGCTCTTCCGCTTAACTTGTCTACTTAAAG	750	
Db	301	AAAAAAGCGGAGAGAGTCCGCTCTTCCGCTTAACTTGTCTACTTAAAG	351	

```

RESULT 3
LOCUS      CL688782
DEFINITION CL688782              791 bp      DNA          linear      GSS 09-JUL-2004
            PRI014b_G04_2 - PRI014b.BR (791) Mixed stage fosmid library of P.
            pacificus var. California Pristionchus pacificus genomic, genomic
            survey sequence.
ACCESSION   CL688782
VERSION     CL688782.1  GI:50197860
KEYWORDS    GSS.
SOURCE      Pristionchus pacificus
            Pristionchus pacificus
            Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
            Neodiplogasteridae; Pristionchus.
            1 (bases 1 to 791)
REFERENCE   Srinivasan,J., Otto,G.W., Kahlow,U., Geisler,R. and Sommer,R.J.
            AppADB: an Acedb database for the nematode satellite organism
            Pristionchus pacificus
JOURNAL     Nucleic Acids Res. 32 (1), D421-D422 (2004)
COMMENT     Contact: Sommer RJ
            Evolutionary Biology
            Max-Planck-Institute for Developmental Biology
            Spemannstr. 37-39, Tuebingen D-72076, Germany
            Tel: 00497071601371
            Fax: 00497071601498
            Email: ralf.sommer@tuebingen.mpg.de
            This library was generated at Caltech, Pasadena, USA and end
            sequenced at Vancouver, Canada.
            Seq primer: T7
            Class: fosmid ends.
FEATURES             location/Qualifiers
             source          1..791
                        /organism="Pristionchus pacificus"
                        /mol_type="genomic DNA"
                        /strain="California"
                        /db_xref="taxon:54126"
                        /clone_lib="Mixed stage fosmid library of P. pacificus
                        var. California"
                        /note="Vector: pEpiFos-5 Fosmid vector"
ORIGIN
Query Match      44.2%;  Score 331.4;  DB 9;  Length 791;
Best Local Similarity 99.7%;  Pred. No. 2.5e-89;
Matches 332;  Conservative 0;  Mismatches 1;  Indels 0;  Gaps 0,

```



Db 58 AGGAAGACGGGCTTCAGGCAAAACCAATCAGCTTTGGGCCAGCTTGCCTTGCAGTT 1

RESULT 6  
BH447440/c  
LOCUS BH447440 819 bp DNA linear GSS 12-DEC-2001  
DEFINITION BOHQB49TF BOHQ Brassica oleracea genomic clone BOHQB49, genomic survey sequence.  
ACCESSION BH447440  
VERSION BH447440.1 GI:17633154  
KEYWORDS GSS.  
SOURCE Brassica oleracea  
ORGANISM Brassica oleracea  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Brassica.  
REFERENCE 1 (bases 1 to 819)  
AUTHORS Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.  
TITLE Whole genome shotgun sequencing of Brassica oleracea  
JOURNAL Unpublished (2001)  
COMMENT Other\_GSSs: BOHQB49TR  
Contact: Chris Town  
TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208  
Email: cdtown@tigr.org  
DNA is from a doubled haploid provided by Tom Osborn.  
Seq primer: TF  
Class: sheared ends.

FEATURES  
source  
1..819  
/organism="Brassica oleracea"  
/mol\_type="genomic DNA"  
/strain="TO1000DH3"  
/db\_xref="taxon:3712"  
/clone="BOHQB49"  
/clone\_lib="BOHQ"  
/note="Vector: PHOS1; Site\_1: BstXI; 2-3 kb sheared genomic DNA inserted into PHOS1 using BstXI linkers"

ORIGIN

Query Match 7.4%; Score 55.2; DB 8; Length 819;  
Best Local Similarity 44.8%; Pred. No. 3.4e-05;  
Matches 256; Conservative 0; Mismatches 313; Indels 3; Gaps 1;

QY 122 CTTTAAAGTGGCTTTTGGACTTACACCCGTGATTACCGCTATGACGCCAGACCGACAAT 181  
Db 734 CCTATATTTTCTTTTGTATTTGTTTATAGTAGCTTCACACCCGACCTTAACAAT 675  
QY 182 ATTCTCGCCCTTAGCTCTGCTACGTCGCATGATTTGTCAAAGTACC CGCGTGGCA 241  
Db 674 TTTTACCGATGACATACGCTAATCAACATGTTAAAAAAGATATGCAATTTGCTT 615  
QY 242 GGGATGAGTCTGGGATTTTGAATGTGATGTACTGTGTGCGGCATTTCAATTTCACTG 301  
Db 614 GGAGTAGCTTTCGGGTTTTCATCCTCACTTCCTATGATGCTTCTTAATATTTCTTA 555  
QY 302 GCAGTGATTGACCCCGGACGGTACACCTTTGAGTTGGCGGGGGCGGCATATATTGTC 361  
Db 554 ATTAAAGTTCTACCCATATGAATTTCCATTAAAAATTTTAAAGGTAGCTTATATGCTC 495  
QY 362 TGGCTGGCGTGAAAAATCGCAACGACCCCAACAAGAGAAGGACTTCAGGCA--AAA 418  
Db 494 TATTTAGCTTTTAAATACTTCAAGTAAAACTAGACACAGATCCCGATGAAAAATATAAT 435  
QY 419 CCAATCAGCTTTTGGCCAGCTTGTCTTGACAGTTTGTGAAGTCAAAATCATTTGTAC 478  
Db 434 AAAAATTATTTACAGTAGGAGATTCTCCTTCAGTTTATTAACCTTAAGGATACATATTC 375  
QY 479 GGTGTTACGGCACTGTGACGTTTGTCTGCGCAACACAGGCGTTAAGCTGGTAGTT 538

Db 374 GGACTAACCGTAGTATCAACTTTTATCTCCCTTATTACAAATCATATTCAGTTATCTA 315

QY 539 GGCGTACGCTTTTGGCTGGCGATGATTGGAGCGTTTGGCAATGTGCTGGCGCTGGCG 598  
Db 314 CTTTTCATTTATTTCTCGGTGATGATTAATGATGATCATGATTTGGAGCTTATTT 255  
QY 599 GGGCATCTGTTTCAGCGATTGTTTCCGACGATATGTCGCCAGTTAAATATCGTGTGCC 658  
Db 254 GGTTCATTTTCAAAACTATATTAACAATATACTAAATCATTTAACAATCATATGACC 195  
QY 659 CTGTGCTGTCATATTGCGCGGTACGCAATTT 690  
Db 194 GTTTATAGTTTGTAGTGCAATTTCAATTGT 163

RESULT 7  
BH374397  
LOCUS BH374397 420 bp DNA linear GSS 10-DEC-2001  
DEFINITION AG-ND-180D20.TR ND-TAM Anopheles gambiae genomic clone  
AG-ND-180D20, genomic survey sequence.  
ACCESSION BH374397  
VERSION BH374397.1 GI:17320539  
KEYWORDS GSS.  
SOURCE Anopheles gambiae (African malaria mosquito)  
ORGANISM Anopheles gambiae  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.  
REFERENCE 1 (bases 1 to 420)  
AUTHORS Hong,Y.S., Hogan,J.R., Wang,X., Sarkar,A., Sim,C., Loftus,B.J., Ren,C., Huff,E.R., Carlile,J.L., Black,K., Zhang,H.-B., Gardner,M.J. and Collins,F.H.  
TITLE Construction of a BAC library and generation of BAC end sequence-tagged connectors for genome sequencing of the African malaria mosquito Anopheles gambiae  
JOURNAL Mol. Genet. Genomics 268 (6), 720-728 (2003)  
MEDLINE 22542063  
PUBMED 12655398  
COMMENT Other GSSs: AG-ND-180D20.TR  
Contact: Brendan J Loftus  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0208  
Fax: 301 838 3543  
Email: bjloftus@tigr.org  
This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center university, College Station, Texas 77843-2123, USA using a HindIII partial digest.  
Seq primer: M13 Rev  
Class: BAC ends.

FEATURES  
source  
1..420  
/organism="Anopheles gambiae"  
/mol\_type="genomic DNA"  
/strain="PEST"  
/db\_xref="taxon:7165"  
/clone="AG-ND-180D20"  
/clone\_lib="ND-TAM"  
/note="Vector: pECBAC1; Site\_1: HindIII"

ORIGIN

Query Match 6.7%; Score 50; DB 8; Length 420;  
Best Local Similarity 49.4%; Pred. No. 0.0011;  
Matches 159; Conservative 0; Mismatches 160; Indels 3; Gaps 1;

QY 147 CCCTGATTACCGCTATGACGCCAGACCGAACAATATTCTCGCCCTTAGCTGTGCTACGT 206



Db 93 CCTTGTAGCTCCGTGACGCGCGCCGCAACAATGATGCTGACGCGCTCGGGGCCA 152  
QY 207 CGCATGATTTGTCMAAGTACCCGCGTGTGGCAGGAGTAGTCTGGATTGTTGATTG 266  
Db 153 CTTTCGGCTACCGCGCATTTGTCCGCATATGCTGGGCATCTGTCTGGGGTGTATCA 212  
QY 267 TGATGTACTGTGTGGCGCATTTCAATTTTCATGGCAGTATGACCCGGCAGCGGTAC 326  
Db 213 TGGTGCTGTATCGCGCGGGCCCTGGCCAGCTGTTCGAGGCCGAACCGCGCATCTACA 272  
QY 327 ACCTTTGAGTTGGCGCGGGCGGCATATATTGTCTGGCTGGCGTGAAATCGCCACCA 386  
Db 273 CCTGCTCAATATATGACAGTGC CGCGTATCTGTCTGGCTGGCGTGAAAGATTGCCGTG 332  
QY 387 GCCCAACAAAGGAAGACGACTTCAGGCAAA--CCAATCAGCTTTTGGGCCACTTTG 443  
Db 333 CCGCAGCGTGATCAGGAGACAGTCAGGCAACCGTCCGCTCAGCTTCTGGCAGCGCCGAG 392  
QY 444 CTTTGACGTTTGTGAAGTCAA 465  
Db 393 CCTTCAGTGGTCAACCCCAA 414

RESULT 8  
LOCUS CL666889 862 bp DNA linear GSS 09-JUL-2004  
DEFINITION PRI0153c\_G02 - PRI0153c.B21 (862) Mixed stage fosmid library of P. pacificus var. California Pristionchus pacificus genomic, genomic survey sequence.  
ACCESSION CL666889  
VERSION CL666889.1 GI:50160241  
KEYWORDS GSS.  
SOURCE Pristionchus pacificus  
ORGANISM Pristionchus pacificus  
REFERENCE Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida; Neodiplogasteridae; Pristionchus.  
AUTHORS 1 (bases 1 to 862)  
TITLESrinivasan,J., Otto,G.W., Kahlow,U., Geisler,R. and Sommer,R.J.  
AppADB: an Acedb database for the nematode satel1ite organism  
Pristionchus pacificus  
Nucleic Acids Res. 32 (1), D421-D422 (2004)  
JOURNAL Contact: Sommer RJ  
COMMENT Evolutionary Biology  
Max-Planck-Institute for Developmental Biology  
Spemannstr. 37-39, Tuebingen D-72076, Germany  
Tel: 00497071601371  
Fax: 00497071601498  
Email: ralf.sommer@tuebingen.mpg.de  
This library was generated at Caltech, Pasadena, USA and end  
sequenced at Vancouver, Canada.  
Seq primer: T7  
Class: fosmid ends.

FEATURES  
source 1..862  
Location/Qualifiers  
1..862  
/organism="Pristionchus pacificus"  
/mol\_type="genomic DNA"  
/strain="California"  
/db\_xref="taxon:54126"  
/clone\_lib="Mixed stage fosmid library of P. pacificus var. California"  
/note="Vector: pBpifos-5 Fosmid vector"

## ORIGIN

Query Match 6.6%; Score 49.4; DB 9; Length 862;  
Best Local Similarity 48.7%; Pred. No. 0.002;  
Matches 134; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

QY 111 TGACACCGACCCCTTTAAGTCTTTTGGACTTACACCCGTGATTACCGCTATGACGCCAG 170  
Db 460 TGTGACCAAGCCTGTATCGGCCCTTGTGCTGTTCCTTCGTTCTTGATCAGCGCCG 401  
QY 171 GACCGAACAATATTCTCGCCCTTAGCTCTGCTACGTGCATGATTTGTCAAAGTACCC 230

Db 400 GTCCGAATACACCATGATCTACTCGCCCTCGGCGCTCAACTTCGGCTTTCCTCGATGC 341  
QY 231 GCGTGTGCAAGGATAGTCTGGATTGTTGATTGTATGTTACTGTGTGCGGCATT 290  
Db 340 CCCATGCGCTGGGATCAGTATCGGCTTCATGCTGCTGTGATTTCGTGCGCCTGGCC 281  
QY 291 CATTTCACTGGCAGTATGATGACCCGGCAGCGGTACACCTTTGAGTTGGCGGGCGG 350  
Db 280 TGGCGAGGTGTTCAGCTATTCCCTTGGGCTTACACGCTGTGTGCGCTATGTGGCGCG 221  
QY 351 CATATATTGTCTGCTGGCGTGGGAAATCGCCACC 385  
Db 220 CGTACCTGCTGTACTGCGGTGAAGATCGCAAC 186

RESULT 9  
LOCUS BH379453 450 bp DNA linear GSS 10-DEC-2001  
DEFINITION AG-ND-163M3.TR ND-TAM Anopheles gambiae genomic clone AG-ND-163M3, genomic survey sequence.  
ACCESSION BH379453  
VERSION BH379453  
KEYWORDS BH379453.1 GI:17325595  
SOURCE GSS.  
ORGANISM Anopheles gambiae (African malaria mosquito)  
Anopheles gambiae  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.

REFERENCE 1 (bases 1 to 450)  
AUTHORS Hong,Y.S., Hogan,J.R., Wang,X., Sarkar,A., Sim,C., Loftus,B.J., Ren,C., Huff,E.R., Carlile,J.L., Black,K., Zhang,H.-B., Gardner,M.J. and Collins,F.H.  
TITLES Construction of a BAC library and generation of BAC end sequence-tagged connectors for genome sequencing of the African malaria mosquito Anopheles gambiae  
Mol. Genet. Genomics 268 (6), 720-728 (2003)  
JOURNAL 12655398  
MEDLINE 22542063  
PUBMED 12655398  
COMMENT Other GSSs: AG-ND-163M3.TF  
Contact: Brendan J Loftus  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0208  
Fax: 301 838 3543  
Email: bjloftus@tigr.org  
This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PBST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.  
Seq primer: M13 Rev  
Class: BAC ends.

FEATURES  
source 1..450  
Location/Qualifiers  
1..450  
/organism="Anopheles gambiae"  
/mol\_type="genomic DNA"  
/strain="PBST"  
/db\_xref="taxon:7165"  
/clone="AG-ND-163M3"  
/clone\_lib="ND-TAM"  
/note="Vector: pECBAC1; Site\_1: HindIII"

## ORIGIN

Query Match 5.8%; Score 43.4; DB 8; Length 450;  
Best Local Similarity 47.9%; Pred. No. 0.12;  
Matches 125; Conservative 0; Mismatches 136; Indels 0; Gaps 0;





AUTHORS	Genoscope.
TITLE	Direct Submission
JOURNAL	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : <a href="mailto:seoref@genoscope.cns.fr">seoref@genoscope.cns.fr</a> )
COMMENT	- Web : <a href="http://www.genoscope.cns.fr">www.genoscope.cns.fr</a> Determination of this BAC-end sequence was carried out as part of a

Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammosser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

## FEATURES

## Bouice

```
1..1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR29104"
/clone_lib="RPCI-98"
note="end : TET3"
```

**ORIGIN**

Query Match 5.1%; Score 38.2; DB 9; Length 1101;  
Best Local Similarity 15.7%; Pred. No. 5.6;  
Matches 43; Conservative 113; Mismatches 118; Indels 0; Gaps 0;

Dy 191 CTTAGCTCTGTAAGTCGCATGGATTTCGTAAAGTACCCGCCGTGTCGCAAGGATGAGT 250  
::|::::: : :::: | || :: : | :: | :: :  
Db 822 MKTKMKMKAMCTKKKMKTKGKGKTTTTDKTMCTTTKTGTGMGKGKTTGTWTTTRMCKT 881

[illegible]

DQ 311 GACCCGCGCAGCGTACACCTTTGAGTTGGGCCGGGCATATATTGTCTGGCTGGCG 370  
::|:: :|::: |::: :||:: : : : :  
Db 942 KAKHCCKTKYKKGKANMGMKDMGKGKDMSKSGKGKBNKNMKMKKKKA CRMKKKKVMC 1001

[illegible]

QY	431	TTGGGCCAGCTTTGCTTGCAGTTTGTGAACGTCA	464
		: ::   : : ::   : : :: : :	
Db	1062	CKAVVMCMCKAKBKKKKACMKAKBKVKBKNC	1095

RESULT 13	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
BM965274/c	BM965274	553 bp mRNA linear EST 20-MAR-2002					
		kj93h10.y1 Ascaris suum female ovary Ascaris suum CDNA 5' similar to contatns element MSRI repetitive element ; , mRNA sequence.					
	BM965274		BM965274	GI:19557341			
		EST.					
		Ascaris suum (pig roundworm)					
		Ascaris suum					

## REFERENCE AUTHORS

1 (bases 1 to 553)  
McCarter, J., Clifton, S., Chiapelli, B., Page, D., Martin, J.,  
Wylie, T., Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B.,  
Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Franklin, C.,  
Tsagarashvili, R., Ronko, I., Kennedy, S., Maguire, L., Beck, C.,  
Underwood, K., Steptoe, M., Allen, M., Person, B., Swaller, T.,

TITLE  
JOURNAL  
COMMENT

Harvey, N., Schurk, R., Kohn, S., Shin, T., Jackson, Y., Cardenas, M.,  
McCann, R., Waterston, R. and Wilson, R.  
The Washington Univ. Nematode EST Project, 1999  
Unpublished (1999)  
Contact: McCarter JF

The Washington Univ. Nematode EST Project, 1999  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
The library was donated by Dr. T. Geary, Pharmacia Inc, Kalamazoo, MI (tgeary@am.pnu.com). DNA sequencing by: Washington University Genome Sequencing Center  
Seq primer: T7 from Gibco  
High quality sequence stop: 485.

## FEATURES

**Source**

```

1. .553
/organism="Ascaris suum"
/mol_type="mRNA"
/db_xref="taxon:6253"
/sex="Female"
/tissue_type="Ovary"
/dev_stage="Adult"
/lab_host="DH10B"
/clone_lib="Ascaris suum female ovary"
/notes="Vector: Lambda Zap II(Stratagene); Site_1: EcoRI;
Site_2: EcoRI; The library was donated by Dr. F. Geary,
Pharmacia Inc, Kalamazoo, MI, (tggeary@am.pnu.com). The
library was created from dissected whole female gonads.
DNA sequencing was done by Washington University Genome
Sequencing Center, St. Louis, MO."

```

# ORIGIN

Query Match	5.0%;	Score 37.6;	DB 5;	Length 553;
Best Local Similarity	47.5%;	Pred. No. 7.2;		
Matches 112;	Conservative	0;	Mismatches 124;	Indels 0;
			Gaps	0;

Oy		446	TTCGAGTTTGTGAACGTCAAATCATTTTGTA CGGTGT TACGGCACTGTCGCGTTTTGYT	505
Dd		334	TTGAAC TTAGGGG AATAAAC CAGC GTTTCGCCCTA ATCTTCCGCAC AGCCC ACTCTTTTTT	275

Qy 506 CTGCCGCAACACAGGCGTTAAGCTGGGTAGTTGGCGCTCAGCGTTTTCCTGGCGATGATT 565

Db 274 TTCCGCGCAAGGATGCTGACCAAGGTGTGCTCTTGAGGGCTTAGCATGATGATG 215

Qy	566	GGACGTTTGGCAATGTGTCTGGGCGCTGGCGGGGCATCTGTTTCAGCCATTGTTTCGC	625
Db	214	GTGGCGTAGGTGATGGCAATGTTTGTCTAGAGAGATGCTGATGCTGGCCGACATTGATGATA	155

QY 626 CAGTATGGTCGCCAGTTAAATATCGTCTTGCCCTGTGCTGCTATATGCGCGGT 681  
||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 154 ATGTTGTCGTAAGAGATGATGATGCTTATGCATGCTGGCTTATGSCCTGGT 99

RESULT 14				
BQ095448/c				
LOCUS				
DEFINITION	BQ095448	568 bp	mRNA	linear EST 08-APR-2002
	kk05d05.y1	Ascaris suum	female ovary	Ascaris suum cDNA 5' similar
	to contains element	MSR1	repetitive element ;	mRNA sequence.
ACCESSION	BQ095448			
VERSION	BQ095448.1	GI:20076665		
KEYWORDS	EST.			
SOURCE	Ascaris suum			
ORGANISM	Ascaris suum (pig roundworm)			

## REFERENCE AUTHORS

1 (bases 1 to 568)

McCarter, J., Clifton, S., Chiapelli, B., Page, D., Martin, J.,  
Wyllie, T., Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B.,  
Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Franklin, C.,  
Tsagarishevili, R., Ronko, I., Kennedy, S., Maguire, L., Beck, C.,  
Underwood, K., Steptoe, M., Allen, M., Person, B., Swaller, T.,





GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2005, 01:36:31 ; Search time 42 Seconds  
(without alignments)  
346.585 Million cell updates/sec

**Title:** US-10-620-487-2

Perfect score: 996  
Sequence: 1 VTPTLSAFWYTYLTITAMP.....RQNIIVLALLVYCAVRIFY 195

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 513545 beqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing:	Minimum Match 0%
	Maximum Match 100%
	Listing first 45 summaries

```
Database :      Issued_Patents_AA:*
1: /cgn2_6/prodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/prodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/prodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/prodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/prodata/1/iaa/PCtUS_COMB.pep:*
6: /cgn2_6/prodata/1/iaa/Backfiles1.pep:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query #			ID	Description
		Match	Length	DB		
1	684	68.7	205	4	US-09-489-039A-12661	Sequence 12661, A
2	272.5	27.4	249	4	US-09-252-991A-25717	Sequence 25717, A
3	250.5	25.2	211	4	US-09-489-039A-10442	Sequence 10442, A
4	232.5	23.3	212	4	US-09-543-681A-4767	Sequence 4767, Ap
5	228.5	22.9	211	4	US-09-543-681A-6708	Sequence 6708, Ap
6	191.5	19.2	238	4	US-09-328-352-7122	Sequence 7122, Ap
7	178	17.9	241	4	US-09-328-352-7159	Sequence 7159, Ap
8	173	17.4	235	4	US-09-252-991A-32455	Sequence 32455, A
9	163.5	16.4	271	4	US-09-328-352-8249	Sequence 8249, Ap
10	162	16.3	214	4	US-09-489-039A-8066	Sequence 8066, Ap
11	158.5	15.6	206	4	US-09-252-991A-18746	Sequence 18746, A
12	155.5	15.6	214	4	US-09-328-352-6025	Sequence 6025, Ap
13	155.5	15.6	237	4	US-09-252-991A-27175	Sequence 27175, A
14	154	15.5	210	4	US-09-328-352-6808	Sequence 6808, Ap
15	153.5	15.4	211	4	US-09-252-991A-20747	Sequence 20747, A
16	153	15.4	214	4	US-09-489-039A-9793	Sequence 9793, Ap
17	151.5	15.2	211	4	US-09-902-540-9913	Sequence 9913, Ap
18	149.5	15.0	250	4	US-09-543-681A-4487	Sequence 4487, Ap
19	148	14.9	211	4	US-09-489-039A-9608	Sequence 9608, Ap
20	146.5	14.7	220	4	US-09-328-352-4903	Sequence 4903, Ap
21	146	14.7	222	4	US-09-328-352-6471	Sequence 6471, Ap
22	138.5	13.9	219	4	US-09-489-039A-7731	Sequence 7731, Ap
23	132	13.3	208	4	US-09-328-352-5111	Sequence 5111, Ap
24	128.5	12.9	228	4	US-09-543-681A-4854	Sequence 4854, Ap
25	128.5	12.9	240	4	US-09-328-352-7537	Sequence 7537, Ap
26	128	12.9	220	4	US-09-543-681A-4545	Sequence 4545, Ap
27	127	12.8	200	4	US-09-328-352-4773	Sequence 4773, Ap

## ALIGNMENTS

28	126	12.7	205	3	US-09-396-357-2	Sequence 2, Appli
29	125.5	12.6	214	4	US-09-328-352-6695	Sequence 6695, Ap
30	123.5	12.4	228	4	US-09-252-991A-23837	Sequence 23837, A
31	123.5	12.4	277	4	US-09-252-991A-23018	Sequence 23018, A
32	122.5	12.3	201	4	US-09-328-352-6409	Sequence 6409, Ap
33	115.5	11.6	260	4	US-09-252-991A-25992	Sequence 25992, A
34	112.5	11.3	249	4	US-09-328-352-5813	Sequence 5813, Ap
35	112	11.2	217	4	US-09-489-039A-8076	Sequence 8076, Ap
36	111.5	11.2	222	4	US-09-252-991A-26644	Sequence 26644, A
37	109.5	11.0	216	4	US-09-252-991A-18666	Sequence 18666, A
38	107.5	10.8	206	4	US-09-328-352-4784	Sequence 4784, Ap
39	105.5	10.6	209	4	US-09-328-352-8225	Sequence 8225, Ap
40	104.5	10.5	210	4	US-09-328-352-4775	Sequence 4775, Ap
41	103.5	10.4	226	4	US-09-252-991A-31610	Sequence 31610, A
42	102.5	10.3	221	4	US-09-328-352-6659	Sequence 6659, Ap
43	99	9.9	213	4	US-09-489-039A-11845	Sequence 11845, A
44	98.5	9.9	138	4	US-09-711-164-445	Sequence 445, App
45	98.5	9.9	269	4	US-09-489-039A-7320	Sequence 7320, Ap

```

RESULT 1
US-09-489-039A-12661
; Sequence 12661, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 12661
; LENGTH: 205
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-12661

```

Query Match	68.7%;	Score 684;	DB 4;	Length 205;
Best Local Similarity	68.4%;	Pred. No. 3.4e-68;		
Matches 132;	Conservative 21;	Mismatches 40;	Indels 0;	Gaps 0;

OY	1 VTPTLISAFWITYTLITAMTPGNNILALSSATSHGFROSTRVLAGMSLGFLIVMLCAGI	60
Dd	10 VTPTLISAFLLTYLTITALTFPGNNILALSSVTSHGRLRSRLVLAGMSVGFIITMICAL	69
OY	61 SFSLSAVIDPAAVHLWSAGAAYIVWLAWKIATSPTKEDGIQAKPISEFWASPALQFVNKY	120
Dd	70 TFSLYELDSRFTLVLGWIGAYAIIWLAWQIAKSKPATGTPEVEPVGFMASTGLQFVNKY	129
OY	121 ILVGVTAISTFVLPTQTALSNVGVSVLLAMIGTFGNVCWALLGHLFORLFRQYGRONT	180
Dd	130 ILYGTALTSTFVLPVTREPWLISVSLLLAIGALGNLCWALLAGHLEFORLFLLYGRQLNW	189
OY	181 VLALLLVYCAVRI	193
Dd	190 MLALLLVYCAVRI	202

RESULT 2  
 US-09-252-991A-25717  
 ; Sequence 25717, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 107196.136



```

; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 25717
; LENGTH: 249
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-25717
```

```

Query Match          27.4%; Score 272.5; DB 4; Length 249;
Best Local Similarity 37.0%; Pred. No. 3.1e-22;
Matches 71; Conservative 39; Mismatches 67; Indels 15; Gaps 8;
```

```

QY 5 LLSAFWTTYTLTAMTPGPNNTLLSSATSHGFRQSTRVLGMSLGFLIVMLLCAGISFSL 64
   ||| : : : ||||| : :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 51 LLLAFSLFAFVTSVTPGPNNTMLLASGVNFGFVRSIPHLIGISCGFFIMWM---AVGFGL 107

QY 65 AVI--DPAVHLLSWAGAYIVWLWKIATS-PTKEDGL-QAKPISFWASPALQFVNVK 119
   : :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 108 GTVEAYPVLYTLRYGAYLLYLWKIATSGPAGNDPEGKPLSYWGAAAFQWVNP 167

QY 120 ILYGVTAISTFVLPTQALSWVGVSVLLAMIG--TEGNVCWALAGHLFQRLFR--QYG 175
   : :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 168 AMWMAVGAISTYT-PLQGYFTNVVISTVFALINAPTIG--IWAGFGSMLRNVLRDRWL 224

QY 176 RQNLIVLALLV 187
   | :| :| :| :| :|
Db 225 RVFNVMALLLV 236
```

```

RESULT 3
US-09-489-039A-10442
; Sequence 10442, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 10442
; LENGTH: 211
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-10442
```

```

Query Match          25.2%; Score 250.5; DB 4; Length 211;
Best Local Similarity 32.4%; Pred. No. 7e-20;
Matches 61; Conservative 42; Mismatches 78; Indels 7; Gaps 5;
```

```

QY 5 LLSAFWTTYTLTAMTPGPNNTLLSSATSHGFRQSTRVLGMSLGFLIVMLLCAGISFSL 64
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 17 LLAFIYAFVTSITPGPNNTMLLASGLNVGFKSLPHVLGISIGFAL-MVIAVGTGLGR 75

QY 65 AVIDPAVHL-LSWAGAYIVWLWKIATS--PTKEDGLQAKPISFWASPALQFVNVI 121
   : :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 76 VFEENPOVYIALRICGATYLLIYLAWNIAATALPMSKISSESSARPFSEWAAGFQWVNP 135

QY 122 LYGVTAISTFVLPTQALSWVGVSVLLAMIGTFCNVCMALAGHLFQRLFR--QYGRQ 179
   : :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 136 IMATGALITTY-LPQGENVRSVIFSLIFASINAPSVSIWTFGATLRHMLNDIKYLRIF 194

QY 180 IVLALLV 187
   : :| :| :| :| :| :|
Db 195 VCMAMLL 202
```

```

RESULT 4
US-09-543-681A-4767
; Sequence 4767, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABI
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 4767
; LENGTH: 212
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-4767
```

```

Query Match          23.3%; Score 232.5; DB 4; Length 212;
Best Local Similarity 30.2%; Pred. No. 7.1e-18;
Matches 57; Conservative 45; Mismatches 80; Indels 7; Gaps 6;
```

```

QY 12 YTLTAMTPGPNNTLLSSATSHGFRQSTRVLGMSLGFLIVMLLCAGISF-SLAVIDPA 70
   : :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 25 FSVTSITPGPNNTMLLASGINFGLKRTMPHAIQVSLGF-FVWLLAVGIGALIKSSPI 83

QY 71 AVHLLSWAGAYIVWLWKIATSPTKE--DGLQAKPISFWASPALQFVNVIILYGVTA 128
   : :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 84 IYNILKYLGALEYLLWLWKTAISHSVEQNSNKGSPFLTLLAALFQWVNPKSMMAISGI 143

QY 129 STFVLPTQALSWVGVSVLLAMIGTFCNVCMALAGH-LFQRLFR-QYGRQNLIVALL 186
   : :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 144 TLYTSPQYPIISMILL-VAIIFTLINPCCVAIWATFGHSRLRRLKNPKILKLFNFIMG 202

QY 187 VYCAVRIFY 195
   | :| :| :| :| :|
Db 203 ALSAISVL 211
```

```

RESULT 5
US-09-543-681A-6708
; Sequence 6708, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABI
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 6708
; LENGTH: 211
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-6708
```

```

Query Match          22.9%; Score 228.5; DB 4; Length 211;
Best Local Similarity 31.5%; Pred. No. 2e-17;
Matches 62; Conservative 43; Mismatches 81; Indels 11; Gaps 7;
```

```

QY 1 VTPTLLSAFWTTYTLTAMTPGPNNTLLSSATSHGFRQSTRVLGMSLGFLIVMLLC-A 59
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 11 MTNSLVSVLSVFLFAIAITPGPNNTLLTSSGAYGKRSILTMAGIIVGQCVLSSAF 70

QY 60 ISFSLAVIDPAVHL-LSWAGAYIVWLWKIATSPTKEDGLQAK--PISFWASPALQ 115
   : :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
```

Db 71 1A-TLLIYP-ALHIGLIGSIYLCWLAWKTATASYORLDISSKVSQTTAFQGLLOF 128  
QY 116 VNVKILYGTALSTFVLPQTQALSWVGVSVLLAMIGTFGNVCWALAGHLFORLPRQYG 175  
Db 129 LNPKAMMGLGAVGSFSLAGDYGISGIVSVMLLVNFIAGMWILGTFISR-FLQSR 187  
QY 176 RQ---LNIYVALLLVYC 189  
Db 188 RAWFLFNIMGILTAMC 204

RESULT 6  
US-09-328-352-7122  
; Sequence 7122, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 7122  
; LENGTH: 238  
; TYPE: PRT  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-7122

Query Match 19.2%; Score 191.5; DB 4; Length 238;  
Best Local Similarity 27.1%; Pred. No. 3.1e-13;  
Matches 52; Conservative 46; Mismatches 79; Indels 15; Gaps 8;  
QY 6 LSAFWTYTLITAMTPGPNNILALSSATSHGFRQSTRVLAGMSLGLIYVMLCAGISFSLA 65  
Db 41 LTAFLFAIVASITPGPTNFILSLSHYKISKITLPEVILGSCIGALLLVV-VGIGLGST 99  
QY 66 VID-PAVHLLSWAGAIYVWLAKIAT-SP--TKEDGLQAKPISFWASFPALQFVNKII 121  
Db 100 ILAYPVIQKIMAMSGLIWTLVLAWKLNNPNVISLEKNEQYPPIGFKAFLMQALNPKTW 159  
QY 122 LYGTALSTFVLPQTQALSWVGVSVLLAMIGTFGNVC---WALAGHLFORLPRQYGRQL 178  
Db 160 MMAFAVISVYT---KQGDILVNVSILSCIFLLIAPFCLYLVALVGRLSTRLSK-PKHI 215  
QY 179 NI---VLALLV 187  
Db 216 NIFNKIMAILL 227

RESULT 7  
US-09-328-352-7159  
; Sequence 7159, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 7159  
; LENGTH: 241  
; TYPE: PRT  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-7159

Query Match 17.9%; Score 178; DB 4; Length 241;  
Best Local Similarity 24.6%; Pred. No. 1e-11;  
Matches 56; Conservative 48; Mismatches 78; Indels 46; Gaps 10;  
QY 1 VTPTLLSAF-----WTYTL-----ITAMTPGPNNILALSSATSHGFRQSTRVL 43

Db 17 MNPTYILHFIPSTRTYTMTAILIPYLAITLLTLTPGLDITLIRATLEGKSKAFQAA 76  
QY 44 AGMSLGLI--VMLLCAGISFSLAVIDPAVHLLSWAGAIYVWLAKIATSPTK----- 96  
Db 77 LGISLGIAMGIYVAC-GLG-ALLMASDIAFNILKMGAAIYLAWLGLNMLIKPRSQLADI 134  
QY 97 EDGLQAKPIS---FWASFPALQFVNKILYGTALSTFVLPQTQALSWVGVSVLLAMIG 153  
Db 135 QDNHSNRSTSENVFIKFFGNLNPVKYGIYISFLPQFIPAQASAVTWVWGLVMIHVIG 194  
QY 154 TEGNVCWA-----LAGHLFORLPRQYGRQLN---IVLALLVY 188  
Db 195 ----VLWSSLILAMQPLSRYLKQPKFVKYMDRITGSIYVLFALKLAF 238

RESULT 8  
US-09-252-991A-32455  
; Sequence 32455, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 32455  
; LENGTH: 271  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-32455

Query Match 17.4%; Score 173; DB 4; Length 271;  
Best Local Similarity 27.9%; Pred. No. 4.3e-11;  
Matches 56; Conservative 36; Mismatches 79; Indels 30; Gaps 7;  
QY 5 LLSAFWYTYTLITAMTPGPNNILALSSATSHGFRQSTRVLAGMSLGLIYVMLCAGISFSL 64  
Db 73 LLLAMGAFSLSLISIPGVNLTIVASGANHGFRRTLPTVTGATLGF-VLLAFVGFWR 131  
QY 65 AV-IDPAVHLLSWAGAIYVWLAKIATSPTK---EDGLQAKPISFWASFPALQFVNK 119  
Db 132 AIEAYPRFFDYLGMAGAFAIHVGYRIATDPRIALLENGVP---GFFQGVLLQWLNP 187  
QY 120 IILYGTALSTFVLPQTQALSWVGVSVLLAMIGTFGNVC-----WALAGHLF----- 167  
Db 188 AWIACASGVALFASPSTHA-----PLIVFMAIYLVVCYLSLAAWALLGDRVALLDS 239  
QY 168 QRLFRQYGRQLNIYVALLVY 188  
Db 240 PRRVRLFNFRAMGGTLVLTAGY 260

RESULT 9  
US-09-328-352-8249  
; Sequence 8249, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 8249  
; LENGTH: 235



```
; Sequence 27175, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27175
; LENGTH: 237
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27175
```

```
Query Match      15.6%; Score 155.5; DB 4; Length 237;
Best Local Similarity 30.4%; Pred. No. 3.2e-09;
Matches 51; Conservative 33; Mismatches 63; Indels 21; Gaps 9;
```

```
OY      3 PTL-LSAFWYTL---ITAMTPGPNNILALSSATSHGFRQSTRVLGMSLGFILVLLCA 58
      |||::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      23 PTLGIDTFWTVLVGVFVLLIPGNSLFLVATSAQRGVAATGYRACGVFLGDVAILMLISA 82

OY      59 -GISFSLAVIDPAVAHLISWAGAAYIVWL-----AWKIATSPTEKEDGLQAKPIS---- 106
      ||::||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      83 LGVA-SLKAEPMFLIGLYLGAAYLFYLGVMRGAMRKLRNP-EATAQAQEKVDVHQP 140

OY      107 FWASFALQFVNVKIILYGVLTALSTFVLQ-QLQALSWVGVSVLLAMI 152
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      141 FRKALLSLSNPKAILFFISFFIQFVDPGYAYPGLSFLV-LAVILLELV 187
```

```
RESULT 14
US-09-328-352-6808
; Sequence 6808, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 6808
; LENGTH: 210
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-6808
```

```
Query Match      15.5%; Score 154; DB 4; Length 210;
Best Local Similarity 23.6%; Pred. No. 3.9e-09;
Matches 48; Conservative 53; Mismatches 78; Indels 24; Gaps 7;
```

```
OY      6 LSAFWTYTLITAMTPGPNNILALSSATSHGFRQSTRVLGMSLG-FLIVMLCAGISFSL 64
      |||::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      9 LLVFITIAFLTLSSPGVLFVTNSINYGVRTALFGISGLITGMFIIVASAGVGL-I 67

OY      65 AVIDPAVHLISWAGAAYIVWLAK--IATSPK-----EDGLQAKPISFWASFALQ 114
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      68 ITSNPTIFTALKFIAFLMYLGYKNFIKKSPTQDLLETHNEKIKSKLFGQGLFA-S 126

OY      115 FVNWKIILYGVLTALSTFVLQQLQALSWVGVSVLLAMIGTFGNVCMA-----LAG 164
      |||::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      127 LNPKTIVFFIALFPQFIDIKKEILNQFLVSLTFLIGFLIHLVYANFSSIFREKMLAG 186

OY      165 HLFQRLFRQYGRQNLIVALLLV 187
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
```

```
Db      187 NNFSKLNKVSQ-CIFELLAVLLI 208

RESULT 15
US-09-252-991A-20747
; Sequence 20747, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 20747
; LENGTH: 211
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-20747
```

```
Query Match      15.4%; Score 153.5; DB 4; Length 211;
Best Local Similarity 28.4%; Pred. No. 4.5e-09;
Matches 50; Conservative 28; Mismatches 75; Indels 23; Gaps 7;
```

```
OY      18 MTPGPNNILALSSATSHGFRQSTRVLGMSLGFILVLLCAGISFSLAVIDPAVAHLISW 77
      :|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      21 LAPGPNNILSNNAARHGFPATASLAGGRLAFAGMLALAAS---GLALV----LHTSAW 73

OY      78 -----AGAAIVWLA---WKIATSPTEKEDGLQAKPISFWASFALQFV----NVKIILY 123
      |||::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      74 LFLAIKVLGAAYLLMLAVQLWRTDAQPLVNEASPARP-SLWRLGRQEFVVAAGNPKAILI 132

OY      124 GVTALSTFVLQQLQALSWVGVSVLLAMIGTFGNVCWALAG-HLFQRLFRQYGRQL 178
      |||::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      133 FTAFLPQFVDPGQPLGAQFAQLGAFFLLEWLAIALYSYAGLHGRLLAGQPARRL 188
```

Search completed: September 10, 2005, 01:46:43  
Job time : 43 secs



**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2005, 01:45:17 ; Search time 168 Seconds  
(without alignments)  
457.822 Million cell updates/sec

Title: US-10-620-487-2

Perfect score: 996

Sequence: 1 VTPTLLSAFWTYTLITAMTP.....RQLNIVLALLLVYCAVRIFY 195

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1777461 seqs, 394431504 residues

Total number of hits satisfying chosen parameters: 1777461

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10D\_PUBCOMB.pep:\*  
17: /cgn2\_6/ptodata/1/pubpaa/US10E\_PUBCOMB.pep:\*  
18: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*  
19: /cgn2\_6/ptodata/1/pubpaa/US11A\_PUBCOMB.pep:\*  
20: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep:\*  
21: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*  
22: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	996	100.0	195	US-10-620-487-2	Sequence 2, Appli
2	161.5	16.2	206	US-10-282-122A-60157	Sequence 60157, A
3	158.5	15.9	211	US-10-282-122A-67883	Sequence 67883, A
4	157.5	15.8	206	US-10-282-122A-56211	Sequence 56211, A
5	156.5	15.7	206	US-10-282-122A-75782	Sequence 75782, A
6	156.5	15.7	206	US-10-282-122A-78024	Sequence 78024, A
7	154	15.5	205	US-10-282-122A-72863	Sequence 72863, A
8	151.5	15.2	211	US-10-282-122A-47826	Sequence 47826, A
9	142	14.3	217	US-10-156-761-10918	Sequence 10918, A
10	137.5	13.8	205	US-10-282-122A-76964	Sequence 76964, A
11	126	12.7	205	US-09-927-395-2	Sequence 2, Appli

12	126	12.7	205	9	US-09-847-392-2	Sequence 2, Appli
13	123.5	12.4	209	15	US-10-282-122A-66814	Sequence 66814, A
14	120	12.0	181	15	US-10-282-122A-44873	Sequence 44873, A
15	117.5	11.8	210	15	US-10-282-122A-69338	Sequence 69338, A
16	116.5	11.7	223	9	US-09-738-626-6070	Sequence 6070, Ap
17	116.5	11.7	223	10	US-09-746-660A-14	Sequence 14, Appl
18	116.5	11.7	223	17	US-10-494-672-296	Sequence 296, App
19	116.5	11.7	223	17	US-10-494-675-108	Sequence 108, App
20	116.5	11.7	232	17	US-10-762-107-86	Sequence 86, Appl
21	112	11.2	207	9	US-09-738-626-6418	Sequence 6418, Ap
22	112	11.2	227	15	US-10-461-194-124	Sequence 124, App
23	111	11.1	227	15	US-10-461-194-94	Sequence 94, Appl
24	110.5	11.1	205	14	US-10-156-761-11297	Sequence 11297, A
25	109.5	11.0	209	15	US-10-282-122A-51102	Sequence 51102, A
26	107.5	10.8	224	14	US-10-156-761-7796	Sequence 7796, Ap
27	104	10.4	226	9	US-09-738-626-3665	Sequence 3665, Ap
28	98.5	9.9	138	14	US-10-287-274-445	Sequence 445, App
29	98.5	9.9	138	15	US-10-282-122A-42680	Sequence 42680, A
30	96	9.6	948	16	US-10-437-963-161288	Sequence 161288, A
31	95.5	9.6	473	15	US-10-282-122A-61289	Sequence 61289, A
32	95	9.5	537	15	US-10-282-122A-68111	Sequence 68111, A
33	93.5	9.4	397	17	US-10-374-903A-6	Sequence 6, Appl1
34	92	9.2	486	15	US-10-369-442-16	Sequence 16, Appl1
35	92	9.2	486	15	US-10-369-493-15301	Sequence 15301, A
36	92	9.2	488	15	US-10-369-493-15667	Sequence 15667, A
37	92	9.2	488	15	US-10-369-493-16057	Sequence 16057, A
38	92	9.2	625	15	US-10-369-493-3850	Sequence 3850, Ap
39	91.5	9.2	264	11	US-09-864-408A-4810	Sequence 4810, Ap
40	91.5	9.2	415	18	US-10-859-210-4	Sequence 4, Appl1
41	91	9.1	246	15	US-10-424-599-253409	Sequence 253409, A
42	91	9.1	470	16	US-10-767-701-44234	Sequence 44234, A
43	91	9.1	551	9	US-09-738-626-4431	Sequence 4431, Ap
44	91	9.1	551	15	US-10-627-476-510	Sequence 510, App
45	90.5	9.1	344	18	US-10-724-972A-7313	Sequence 7313, Ap

ALIGNMENTS

RESULT 1  
US-10-620-487-2  
; Sequence 2, Application US/10620487  
; Publication No. US20040038352A1  
; GENERAL INFORMATION:  
; APPLICANT: MAIER, THOMAS  
; TITLE OF INVENTION: METHOD FOR FERMENTATIVE PRODUCTION OF AMINO ACIDS AND AMINO ACID  
; FILE REFERENCE: MAIER, T-2  
; CURRENT APPLICATION NUMBER: US/10/620,487  
; PRIOR FILING DATE: 2003-07-16  
; PRIOR APPLICATION NUMBER: GERMAN NO. 102 32 930.3  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2  
; LENGTH: 195  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-10-620-487-2

Query Match	100.0%;	Score 996;	DB 15;	Length 195;
Best Local Similarity	100.0%;	Pred. No. 2.6e-92;		
Matches 195;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	VTPTLLSAFWTYTLITAMTPGPNNTILSSATSHGFRQSTRVLAKMSLGLFLIVMLLCAGI	60	
Db	1	VTPTLLSAFWTYTLITAMTPGPNNTILSSATSHGFRQSTRVLAKMSLGLFLIVMLLCAGI	60	
QY	61	SFSLAVIDPAAVHLISWGAAYIWLAKIATSPTEKEDGLQAKPIISFWASPALQFVNVKI	120	
Db	61	SFSLAVIDPAAVHLISWGAAYIWLAKIATSPTEKEDGLQAKPIISFWASPALQFVNVKI	120	
QY	121	ILYGVTLSTFVLPGTQALSWVVGVSVLAMIGTFGNVCWALAGHLFQRLFRQYGRQLNI	180	

```

Db      121 ILYGTALSTFVLPTOTALSWVGVSVLLAMIGTGNVCWALAGHLFQRLFRÖYGRÖLNI 180
QY      181 VLALLLVYCAVRIFY 195
Db      181 VLALLLVYCAVRIFY 195
```

RESULT 2

```

US-10-282-122A-60157
; Sequence 60157, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 60157
; LENGTH: 206
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-10-282-122A-60157
```

Query Match 16.2%; Score 161.5; DB 15; Length 206;  
Best Local Similarity 27.0%; Pred. No. 5e-08;  
Matches 53; Conservative 41; Mismatches 81; Indels 21; Gaps 9;

```

QY      8 AFWTYTLITAMTPGPNNILALSSATSHGFRQSTRVLAGMSLGFLIVMLLCA---GISFSL 64
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      8 AYLTSTIILSLSPSGAINTMTTISINHGVRGAASIAIGLOTGLVIHIVLVGVGLTFSR 67
QY      65 AVIDPAVHLLSWAGAAYIVWLA---WKIATSPTKED--GIQAKPISFWASFALQFVNVK 119
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      68 SVL---AFEVLKWAGAAYLWLGIQWRAGAIDNLTLAKAQTRGKLFQRAVFNLTNPK 124
QY      120 ILYGVNTALSTFVLPTOTALSWVGVSV---LLAMIG--TFGN--VCWALAGHLFQR 169
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      125 SIPELAALFPQFILPHQFQWQVQLVLTGTTIVVDIIVMIGVATLAQRISAWIKGPQMK 184
```

```

QY      170 LFRÖYGRÖLNIVLALL 185
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      185 LNKVFGSLFMLVGALL 200
```

RESULT 3

```

US-10-282-122A-67883
; Sequence 67883, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 67883
; LENGTH: 211
; TYPE: PRT
; ORGANISM: Pseudomonas putida
US-10-282-122A-67883
```

Query Match 15.9%; Score 158.5; DB 15; Length 211;  
Best Local Similarity 26.3%; Pred. No. 1e-07;  
Matches 52; Conservative 41; Mismatches 82; Indels 23; Gaps 9;

```

QY      9 FWTYTLITAMTPGPNNILALSSATSHGFRQSTRVLAGMSLGFLI-VMLLCAGISFSLAVI 67
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      10 FFAACWVISLSFGAGAIASMSGLQYGFWRGYWNALGLQLGIMQIAIAAGVAVLAA- 68
QY      68 DPAVHLLSWAGAAYIVWLAWK---IATSPTKEDGLQ--AKPISF-WASFALQFVNVKI 120
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      69 SATAFQVTKMFGVGLVYLAQKWRALPMDMSDESGVRPIGKPLSVRGLVNISNPKA 128
QY      121 ILYGVNTALSTFVLPTOTALSWVGVSV-----LLAMIGTFGNVCWALAGHLFQRLFR-- 172
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      129 LVFMLAVLPQFLNPHAPLLPQYVAITVTWTVVDLIVMAGYTG-----LASHVL-RMLRTP 182
QY      173 QYGRÖLNIVLALLVYCA 190
```

Db : : : : :  
183 KQKRLNRTFAGLFIGAA 200

RESULT 4

```
US-10-282-122A-56211
; Sequence 56211, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See file wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 56211
; LENGTH: 206
; TYPE: PRT
; ORGANISM: Enterobacter cloacae
US-10-282-122A-56211
```

Query Match 15.8%; Score 157.5; DB 15; Length 206;  
Best Local Similarity 27.4%; Pred. No. 1.3e-07;  
Matches 55; Conservative 38; Mismatches 77; Indels 31; Gaps 10;

```
QY 8 AFWTYTLITAMTPGPNNILALSSATSHGFRQSTRVLAGMSLGLIIVMLLCA---GISFSL 64
   | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 8 AYLTSTILSLSPGSGAINTMTTTSINHGVRGAASIAGLOTGLGIHIVLVGIGITLFSR 67
   | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 65 AVIDPAVHLWSAGAAIYVLA--WKIATS-----PTKEDGLQAKPISTWASPALQ 114
   | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 68 SVL--AFEVLKWAAGAYLIWLGIQWRRAAGSINLNTLALTONRGHLFKRAVF-----VN 119
   | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 115 FVNVKIILYGVTLSTFVLV-QTQALSWV-GVSV----LLAMIG--TFGN--VCWALAG 164
   | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 120 LTNPKSIVFLAALFPQFIIVPHQPOVMQYVVLGATTIIVDIIVMIGYATLAQRIAAWIKGP 179
   | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 165 HLFQRLFRQYGRQNLIVLALL 185
   | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 180 KQMKALNKVFGSLFMLVGALL 200
   | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
```

RESULT 5

```
US-10-282-122A-75782
; Sequence 75782, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See file wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 75782
; LENGTH: 206
; TYPE: PRT
; ORGANISM: Salmonella typhi
US-10-282-122A-75782
```

Query Match 15.7%; Score 156.5; DB 15; Length 206;  
Best Local Similarity 26.4%; Pred. No. 1.6e-07;  
Matches 55; Conservative 38; Mismatches 74; Indels 41; Gaps 9;

```
QY 1 VTPTLSAFWTYTLITAMTPGPNNILALSSATSHGFRQSTRVLAGMSLGLIIVMLLCA-- 58
   | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 11 LSTLTS-----LSPGSGAINTMTTTSINHGVRGAASIAGLOTGLGIHIVLVGVG 60
   | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 59 -GISFSLAVIDPAVHLWSAGAAIYVLA--WKIA-----TSPTEKEDGLQAKPISF 107
   | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 61 LGTLFSRSLI---AFELIKWAAGAYLIWLGIQWRRAAGAIIDLHTLAQTOSRGRLLFKRAIF 117
   | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 108 WASPALQFVNVKIILYGVTLSTFVLPTQALS--WVGV-SVLLAMIGTFGN----- 157
   | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 118 -----VNLTPKSIVFLAALFPQFIIVPHQPOQLAQYLLGVTTIIVDMIVMTGYATLAQRI 172
   | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 158 VCWALAGHLFQRLFRQYGRQNLIVLALL 185
   | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 173 AAWIKGPQMKALNKAFGSLFMLVGALL 200
   | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
```



```
RESULT 6
US-10-282-122A-78024
; Sequence 78024, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 78024
; LENGTH: 206
; TYPE: PRT
; ORGANISM: Yersinia pestis
US-10-282-122A-78024

Query Match      15.7%; Score 156.5; DB 15; Length 206;
Best Local Similarity 28.4%; Pred. No. 1.6e-07;
Matches 56; Conservative 31; Mismatches 89; Indels 21; Gaps 9;

QY      9 FW-TY--TLITAMTPGPNNILALSSATSHGFRQSTRVLAGMSLGFLLIVMLCAGISFSL 64
      :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB      5 WWLTYLLTLLSLSPGSAINTWSTAISHGTRGVVASIGGLQLG-LAVHIVLVGVLGA 63
QY      65 AVIDP-AAVHLLSWAGAAIYVWLA--WKIATSPTKEDGLQAKPIS--FWASFALQFVNV 118
      |:|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB      64 LVSQSLAFELIKWLGAAYLIWLGIOQWRAAGSLDLHALANSMPRKLFKRAVFVNLTNP 123
QY      119 KIILYGTALSTFVL-P-QTQALSWV--GVSVLAMIGTFGNVC-----WALAGHLFQ 168
      |:|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB      124 KSIIVFLALFPQFVLPQGPQVACYLLILGSTSVIVDLIVMIGYATLATRIARWIKSPQMK 183
QY      169 RLFRQYGRQNLIVLALL 185
      |:|||:||||:||||:||||:||||:||||:||||:||||:||||:
DB      184 LNRIFGGLFMLIGALL 200
```

RESULT 7
US-10-282-122A-72863

```
; Sequence 72863, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 72863
; LENGTH: 205
; TYPE: PRT
; ORGANISM: Salmonella paratyphi A
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (169)..(169)
; OTHER INFORMATION: X=any amino acid
US-10-282-122A-72863

Query Match      15.5%; Score 154; DB 15; Length 205;
Best Local Similarity 25.6%; Pred. No. 2.9e-07;
Matches 53; Conservative 39; Mismatches 75; Indels 40; Gaps 9;

QY      1 VPTLLSAFWTYTLITAMTPGPNNILALSSATSHGFRQSTRVLAGMSLGFLLIVMLCA-- 58
      :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB      11 LTSTLLS-----LSPGSAINTWTTINHGVRGAASIAIGLTGIGHIVLVGVG 60
QY      59 -GISFLAVIDPAVHLLSWAGAAIYVWLA--WKIAT-----SPTKEDGLQAKPISFW 108
      |:|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB      61 LGTLFSRSL---AFEILKWAGAAIYIWLGIQWRAGAIDLHTLAQTQSRGRLEFKRAIF- 116
QY      109 ASFALQFVNVKILYGTALSTFVL-PQTQALS--WVVGV-SVLAMIGTFG-----NV 158
      |:|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB      117 ----VNLTPKSIYFLALFPQFIMPQGPQLAQYLLIGVTTIVDMIVMTGYATLAXRTA 172
QY      159 CWALAGHLFQRLFRQYGRQNLIVLALL 185
      |:|||:||||:||||:||||:||||:||||:||||:||||:||||:
DB      173 AWIKGPKQMKALNKAFGSLFMLVGALL 199
```



```
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 76964
; LENGTH: 205
; TYPE: PRT
; ORGANISM: Vibrio cholerae
; US-10-282-122A-76964
```

```
Query Match      13.8%; Score 137.5; DB 15; Length 205;
Best Local Similarity 28.0%; Pred. No. 1.3e-05;
Matches 53; Conservative 30; Mismatches 77; Indels 29; Gaps 11;
```

```
QY      10 WTYTLTA---MTPGPNNILALSSATSHGFRQSTRVLGMSLGLFLIVMLCAGISFSLA 65
      |  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|
Db      6 WLAVLTAVVFLAPGSGTWNISNGLSYGRHSIGAIIGLQIG-LACHIVLVGIGIGAL 64

QY      66 VIDPA-AVHLISWAGAYIVWLA---WK---IATSPTKEDGLQAKPISFWASFALQFVN 117
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      65 VAQSALAFLLIKWIGAYLVWLGIQWRDRAPLTATTSHELSQLLR--KAVLINLTN 122

QY      118 VKIILYGTALSTFVLPTQALSW---VVGVS-VLLAMIGTFGNVCWALAGHLFQRLFR 172
      |  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|
Db      123 PKSIVFLVALFPQFIDPFRD--HWPQFLVLGITVTITDAIVMFGYT--ALAAQL-----G 173

QY      173 QYGRQLNIV 181
      :|  |  |  |  |  |
Db      174 RYIRSPNIM 182
```

```
RESULT 11
US-09-927-395-2
; Sequence 2, Application US/09927395
; Patent No. US20020058314A1
; GENERAL INFORMATION:
; APPLICANT: LIVSHITS, VITALY ARKADIEVICH
; APPLICANT: ZAKATAEVA, NATALIYA PAVLOVNA
; APPLICANT: ALCOSHIN, VLADIMIR VENTYAMIOVICH
; APPLICANT: BELAREOVA, ALL VALENTINOVNA
; APPLICANT: TOKHMAKOVA, IRINA LVOVNA
; TITLE OF INVENTION: DNA CODING FOR PROTEIN WHICH CONFERS ON BACTERIUM
; TITLE OF INVENTION: ESCHERICHIA COLI RESISTANCE TO L0-HOMOSERINE AND METHOD
; FILE REFERENCE: 0010-1039-0
; CURRENT APPLICATION NUMBER: US/09/927,395
; CURRENT FILING DATE: 2001-08-13
; PRIOR APPLICATION NUMBER: 09/396,357
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: RU98118425
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 205
; TYPE: PRT
; ORGANISM: Escherichia coli
; US-09-927-395-2
```

Query Match 12.7%; Score 126; DB 9; Length 205;

```
Best Local Similarity 24.5%; Pred. No. 0.00019;
Matches 50; Conservative 36; Mismatches 80; Indels 38; Gaps 10;

QY      8 AFWTYTLITAMTPGPNNILALSSATSHGF-----RQSTRVLGMSLGLFLIVML 56
      |  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|
Db      8 AYLLTSIILLTSPGSAINTMTTSLNHGYPAGGVYCWASDRTGDSYCAGWR----- 58

QY      57 CAGISFSLAVIDPAVHLISWAGAYIVWLA---WKIATSPTKED--GLQAKPISFWASF 111
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      59 GVTLFSSRSVI--AFEVLLKWAGAYLIWLGIQWRRAAGAILDKSLASTQSRRLHFQRAV 115

QY      112 ALQFVNVKIIILYGTALSTFVLPTQALSWV--GVSV---LLAMIG--TFGN--VCWA 161
      :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|
Db      116 FVNLTNPKSIVFLALFPQFIMPQQLMQYIVLVGTTIVVDIIVMIGYATLAQRIALWI 175

QY      162 LAGHLFQRLFRQYGRQLNIVLALL 185
      :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|
Db      176 KGPQKMKALNKIFGSLFMLVGALL 199
```

```
RESULT 12
US-09-847-392-2
; Sequence 2, Application US/09847392
; Patent No. US20020102670A1
; GENERAL INFORMATION:
; APPLICANT: LIVSHITS, VITALY ARKADIEVICH
; APPLICANT: ZAKATAEVA, NATALIYA PAVLOVNA
; APPLICANT: ALCOSHIN, VLADIMIR VENTYAMIOVICH
; APPLICANT: BELAREOVA, ALL VALENTINOVNA
; APPLICANT: TOKHMAKOVA, IRINA LVOVNA
; TITLE OF INVENTION: DNA CODING FOR PROTEIN WHICH CONFERS ON BACTERIUM
; TITLE OF INVENTION: ESCHERICHIA COLI RESISTANCE TO L0-HOMOSERINE AND METHOD
; FILE REFERENCE: 0010-1039-0
; CURRENT APPLICATION NUMBER: US/09/847,392
; CURRENT FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: 09/396,357
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: RU98118425
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 205
; TYPE: PRT
; ORGANISM: Escherichia coli
; US-09-847-392-2
```

```
Query Match      12.7%; Score 126; DB 9; Length 205;
Best Local Similarity 24.5%; Pred. No. 0.00019;
Matches 50; Conservative 36; Mismatches 80; Indels 38; Gaps 10;
```

```
QY      8 AFWTYTLITAMTPGPNNILALSSATSHGF-----RQSTRVLGMSLGLFLIVML 56
      |  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|
Db      8 AYLLTSIILLTSPGSAINTMTTSLNHGYPAGGVYCWASDRTGDSYCAGWR----- 58

QY      57 CAGISFSLAVIDPAVHLISWAGAYIVWLA---WKIATSPTKED--GLQAKPISFWASF 111
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      59 GVTLFSSRSVI--AFEVLLKWAGAYLIWLGIQWRRAAGAILDKSLASTQSRRLHFQRAV 115

QY      112 ALQFVNVKIIILYGTALSTFVLPTQALSWV--GVSV---LLAMIG--TFGN--VCWA 161
      :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|
Db      116 FVNLTNPKSIVFLALFPQFIMPQQLMQYIVLVGTTIVVDIIVMIGYATLAQRIALWI 175

QY      162 LAGHLFQRLFRQYGRQLNIVLALL 185
      :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|
Db      176 KGPQKMKALNKIFGSLFMLVGALL 199
```

```
RESULT 13
US-10-282-122A-66814
; Sequence 66814, Application US/10282122A
; Publication No. US20040029129A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2003-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 66814
; LENGTH: 209
; TYPE: PRY
; ORGANISM: Pseudomonas aeruginosa
US-10-282-122A-66814

Query Match      12.4%; Score 123.5; DB 15; Length 209;
Best Local Similarity 24.1%; Pred. No. 0.00035;
Matches 47; Conservative 34; Mismatches 97; Indels 17; Gaps 6;

Qy      8 AFWTYTLTAMTPGPNNILALSSATSHGFRQSTRVLAGSLGF-LIVMLCAGISPSLAV 66
       ||:   ::|||::||:  :||:  |::|:  |:  ::||:  ||
Db      8 AFFLACWAISLSFGAGAIASMSCGLQYGFARGYWNALGIQLALQIAIVAAGVGALLAT 67
Qy      67 IDPAVHLLSWAGAAYIWMLA---WKIATSPTKEDGLQ--AKPISF-WASFALOEVNVKI 120
       |  |::|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      68 -SALAFLIKWFGVAYLYLVLAVRQWQAPPQALSTDEGRPLRPLTVLRGFLVNASNKA 126
Qy      121 ILYGTALSTFVLPOQTALSW-----VGVSVYLAMIIGTFGNVCWALAGHLFORLF 171
       :::  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      127 VIFMLAVLPQFIDPHQPLLQAQLIMGMTIVVDLIWAGYTGLEARVLRVLRSPROKLIV 186
Qy      172 RQYGRQLNIVLALL 186
Db      187 NRTFASLFVGAAGLL 201

RESULT 14
US-10-282-122A-44873
; Sequence 44873, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
```

```

; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 44873
; LENGTH: 181
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-10-282-122A-44873

Query Match          12.0%; Score 120; DB 15; Length 181;
Best local Similarity 26.8%; Pred. No. 0.00067;
Matches 48; Conservative 35; Mismatches 74; Indels 22; Gaps 9;

QY      28 LSSATSHGFRQSTRVLAGMSLGFLI-VMLLCAGISFSLAVIDPAVHLLSWAGAAYIVWL 86
      :||::|||:|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db       1 MSSGLNYGFRHGYWNALIGLQIALLIQIMIVAAGVGLFAT-TPLAFQAVKMFVAYLLYL 59

QY      87 AWKIATSPFKEDGLQ---AKPIS--FWASPALQFVNVKIILYGVTAALSTFV-LPQTQAL 139
      |:|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db       60 AYLQWTAPPKDIEIQHEKKDKSVSALLNGFVNINSPKAIYFLLAVLPQFLDLSKPQWI 119

QY      140 SWVGVSV-----LlAMIGTFGNVCMAAGHLFQRLFRQYGRQ--LNIIVALLLVYCAV 191
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db       120 QYLIMAAIYVTIDLIWMAGYTG-----LASKVL-RLLRSPKQOKYINRGFAVMFSCAAL 172

RESULT 15
US-10-282-122A-69338
; Sequence 69338, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel

```



```

; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 69338
; LENGTH: 210
; TYPE: PRF
; ORGANISM: Pseudomonas syringae
US-10-282-122A-69338

```

Query Match	11.8%	Score 117.5;	DB 15;	Length 210;
Best Local Similarity	22.6%;	Pred. No. 0.0014;		
Matches 45;	Conservative 40;	Mismatches 89;	Indels 25;	Gaps 9;

QY	9	FWTYTLITAMTPGPNNIIALSSATSHGFROSTRVLAGMSLGFLI-VMLICAGISFSLAVI	67
Db	9	FFAACWIISLSPGAGAIASMSCGLQYGFLRGYNALGLQIALVAQIAIVAAGLGAVLAAS	68
QY	68	DPAVHLTWSAGAAYIVWLA---WKIATSPTK-EDGLQAKPIS-----FWASFALQFVNVI	118
Db	69	E-MAFTLIKMFEGVAYLVYLGIKQMR--ASPTDLADESAVRPVGKPMTLVERGFLVNISNP	125
QY	119	KIILGYVTALSTFVLPTQTALSWMVGVSVLLAMIGTFGNVCWALAGH--LFQRLLR----	172
Db	126	KALIFILAILPFQIEFTAPLEFMQYVIAATMVVDLI-----VMAGYTGLASKVLRAIKT	180
QY	173	-QYGRQNLIVLALLLVYCA	190
Db	181	PROQRRLNRTEFASLFEVGAA	199

Search completed: September 10, 2005, 01:59:10  
Job time : 170 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2005, 01:35:41 ; Search time 40 Seconds  
(without alignments)  
469.057 Million cell updates/sec

Title: US-10-620-487-2

Perfect score: 996

Sequence: 1 VTPTLLSAFWTYTLITAMTP.....ROLNIVLALLVYCAVRIFY 195

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_79: \*  
1: pir1: \*  
2: pir2: \*  
3: pir3: \*  
4: pir4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	993	99.7	195	2	A65036	hypothetical 21.2
2	979	98.3	195	2	D91059	hypothetical prote
3	979	98.3	195	2	B85904	hypothetical prote
4	884	88.8	195	2	AF0830	probable membrane
5	272.5	27.4	205	2	B83356	conserved hypothet
6	235.5	23.6	200	2	A10022	probable ABC-trans
7	172	17.3	204	2	B82166	conserved hypothet
8	169	17.0	222	2	E87264	efflux protein, Ly
9	165.5	16.6	206	2	B91223	homoserine/homoser
10	164	16.5	205	2	A12614	RhtB family transp
11	164	16.5	224	2	H97396	dihydropicolinat
12	163	16.4	210	2	E87252	efflux protein, Ly
13	162.5	16.3	249	2	AG3485	homoserine/homoser
14	160.5	16.1	222	2	F82353	conserved hypothet
15	158.5	15.9	197	2	B83280	hypothetical prote
16	158.5	15.9	216	2	C55580	hypothetical prote
17	158	15.9	210	2	D84016	hypothetical prote
18	156.5	15.7	206	2	AF0917	homoserine/homoser
19	156.5	15.7	206	2	AG0466	probable homoserin
20	156.5	15.7	235	2	AD2904	RhtB family transp
21	156.5	15.7	235	2	F97679	hypothetical prote
22	155.5	15.6	209	2	E82388	conserved hypothet
23	155.5	15.6	216	2	F83051	conserved hypothet
24	154.5	15.5	207	2	D83187	hypothetical prote
25	153	15.4	212	2	H87498	lysE family transp
26	143.5	14.4	212	2	F64940	hypothetical prote
27	143	14.4	212	2	AF2805	homoserine/homoser
28	143	14.4	273	2	F97584	hypothetical prote
29	142.5	14.3	212	2	G85790	hypothetical prote

30	140.5	14.1	212	2	C90942	hypothetical prote
31	139	14.0	212	2	AD0714	probable membrane
32	137.5	13.8	205	2	G82358	conserved hypothet
33	135.5	13.6	278	2	E98301	hypothetical prote
34	134.5	13.5	210	2	AC2982	RhtB family transp
35	130	13.1	207	2	E83703	hypothetical prote
36	129	13.0	208	2	C82471	conserved hypothet
37	123.5	12.4	206	2	C82979	hypothetical prote
38	123.5	12.4	209	2	C82990	hypothetical prote
39	123	12.3	213	2	C82523	amino acid transpo
40	122.5	12.3	205	2	AH3641	transporter, lysE
41	122	12.2	206	2	C82139	conserved hypothet
42	122	12.2	216	2	A97675	amino acid efflux-
43	120	12.0	208	2	G87305	efflux protein, Ly
44	119.5	12.0	204	2	B82410	conserved hypothet
45	119.5	12.0	208	2	G84086	dihydropicolinat

ALIGNMENTS

RESULT 1  
A65036  
hypothetical 21.2 kD protein in srmb-ung intergenic region - Escherichia coli (strain K  
C/Species: Escherichia coli  
C/Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 09-Jul-2004  
C/Accession: A65036  
R/Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C  
.A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A/Title: The complete genome sequence of Escherichia coli K-12.  
A/Reference number: A64720; MUID:97426617; PMID:9278503  
A/Accession: A65036  
A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-195 <BLAT>  
A/Cross-references: UNIPROT:P38101; GB:AE000344; GB:U00096; MID:g1788927; PIDN:AACT5631  
A/Experimental source: strain K-12, substrain MG1655  
C/Genetics:  
A/Gene: yfiK

Query Match 99.7%; Score 993; DB 2; Length 195;  
Best Local Similarity 99.5%; Pred. No. 1.9e-81;  
Matches 194; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VTPTLLSAFWTYTLITAMTPGPNNILASATSHGFRQSTRVLAGMSGLIVMLCAGI 60  
:|||||  
Db 1 MTPTLLSAFWTYTLITAMTPGPNNILASATSHGFRQSTRVLAGMSGLIVMLCAGI 60

QY 61 SFSIAVIDPAVHLLSWAGAIYIWLAMKIATSPTEKEDGLQAKPISFNASFAIQFVNKI 120  
|||||  
Db 61 SFSIAVIDPAVHLLSWAGAIYIWLAMKIATSPTEKEDGLQAKPISFNASFAIQFVNKI 120

QY 121 ILGYVTALSTFVLPQTQALSWVGVSVLLAMIGTFGNVCWALAGHLFQRLFRQYGRQLNI 180  
|||||  
Db 121 ILGYVTALSTFVLPQTQALSWVGVSVLLAMIGTFGNVCWALAGHLFQRLFRQYGRQLNI 180

QY 181 VLALLLVYCAVRIFY 195  
|||||  
Db 181 VLALLLVYCAVRIFY 195

RESULT 2  
D91059  
hypothetical protein ECs3444 [imported] - Escherichia coli (strain O157:H7, substrain R  
C/Species: Escherichia coli  
C/Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 09-Jul-2004  
C/Accession: D91059  
R/Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G  
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A/Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen  
A/Reference number: A99629; MUID:21156231; PMID:11258796

A;Accession: D91059  
A;Status: preliminary  
A:Molecule type: DNA  
A;Residues: 1-195 <HAY>  
A;Cross-references: UNIPROT:Q8XA19; GB:BA000007; PIDN:BAB36867.1; PID:g13362915; GSPDB:C  
A;Experimental source: strain O157:H7, substrain RIMD 0509952  
C:Genetics:  
A;Gene: ECs3444

Query Match	98.3%;	Score 979;	DB 2;	length 195;
Best Local Similarity	96.9%;	Pred. No. 3.3e-80;		
Matches 189;	Conservative 4;	Mismatches 2;	Indels 0;	Gaps 0;
QY	1	VTPTLLSAFWTYTLITAMTEGPNNILALSSATSHGFRQSTRVLAGMSLGFLIWMLLCAGI	60	
	:			
Db	1	MTPTLLSAFWTYTLITAMTEGPNNILALSSATTHGHQSTRVLAGMSLGFLIWMLLCAGI	60	
QY	61	SFSLAVIDPAAVHLTSMWGAAYIVWLAMKIATSPTEKEDGLOAKPISFWASFALQFVNWKI	120	
	:			
Db	61	SFSLAVIDPAAVHLTSMWGAAYIVWLAMKIATSPTEKEDGLOTKPISFWASFALQFVNWKI	120	
QY	121	ILYGTALSTFVLPTQTALSMWVGVSVLLAMIGTFGNVCNALAGHLFQRLFRQYGRQLNI	180	
	:			
Db	121	ILYGTALSTFVLPTQTALSMIVGVSVLLAMIGTFGNVCNALAGHLFQRLFRQYGRQLNI	180	
QY	181	VLALLLVYCAVRIFY 195		
	:			
Db	181	VLALLLVYCAVRIFY 195		

RESULT 3  
B85904  
hypothetical protein yfik [imported] - Escherichia coli (strain O157:H7, substrain EDL93  
C/Species: Escherichia coli  
C/Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 09-Jul-2004  
C/Accession: B85904  
R/Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
iller, L.; Grotbeck, E.U.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,  
Nature 409, 529-533, 2001  
A/Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A/Reference number: A85480; MUID:21074935; PMID:11206551  
A/Accession: B85904  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-195 <STO>  
A/Cross-references: UNIPROT:Q8XA19; GB:AE005174; NID:g12516999; PIDN:AAG57694.1; GSPDB:G  
A/Experimental source: strain O157:H7, substrain EDL933  
C/Genetics:  
A/Gene: yfik

[illegible]

RESULT 4  
AF0830

probable membrane protein STY2838 [imported] - Salmonella enterica subsp. enterica sero  
C;/Species: Salmonella enterica subsp. enterica serovar Typhi  
A;/Note: this species has also been called Salmonella typhi  
C;/Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
C;/Accession: AF0830  
R;/Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher  
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar  
, S.; Moule, S.; O'Gaora, P.  
Nature 413, 848-852, 2001  
A;/Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.  
A;/Title: Complete genome sequence of a multiple drug resistant Salmonella enterica sero  
A;/Reference number: AB0502; MUID:21534947; PMID:11677608  
A;/Accession: AF0830  
A;/Status: preliminary  
A;/Molecule type: DNA  
A;/Residues: 1-195 <PAR>  
A;/Cross-references: GB:AL513382; PIDN:CAD02794.1; PID:gl6503803; GSPDB:GN00176  
C;/Genetics:  
A;/Gene: STY2838

	Query Match	88.8%;	Score 884;	DB 2;	Length 195;	
	Best Local Similarity	83.6%;	Pred. No. 9.9e-72;			
	Matches 163;	Conservative 19;	Mismatches 13;	Indels 0;	Gaps 0;	
QY	1 VTPTLLSAFWTYTLTAMTPGPNNILALSSATSHGFROSTRYLAGMSLGFLIVMLLCAGI	60				
Db	1 MTPMLLSAFWTYTLTALTLPGPNNILALSAATAHGFROSIRVLAGMSLGLFVVMLCAGI	60				
QY	61 SFSLAVIDPAAVHLISWAGAAYIVWLAWKIATSPTRKEDGIQAAPISFWASFALOFPVNWKI	120				
Db	61 AFSLAVIDPAIHLSWGAAYITLMAWKIATSPPADEKVRPKPEGVFWVSFGLOFPVNWKI	120				
QY	121 ILYGVTALSTFVLPGTQALSMWVGSVLLAMIGTFGNVCALLAHHLFORLFRQYGRQLNI	180				
Db	121 ILYGITALSTFVLPQTQALNMVITGVSIILLALIGTFGNVCALLAHHLFORAFRHYGRQLNI	180				
QY	181 VLALLLVYCAVRIFY 195					
Db	181 ILALLLVYCAVRIFY 195					

RESULT 5  
B83356  
conserved hypothetical protein PA2306 [imported] - *Pseudomonas aeruginosa* (strain PA01)  
C/Species: *Pseudomonas aeruginosa*  
C/Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 09-Jul-2004  
C/Accession: B83356  
R/Stover, C.K.; Pham, X.Q.; Brwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim  
., Loxy, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A/Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic path  
A/Reference number: A82950; WUID:20437337; PMID:10984043  
A/Accession: B83356  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-205 <STO>  
A/Cross-references: UNIPROT:Q911G9; GB:AE004657; GB:AE004091; NID:g9948338; PIDN:AAG056  
A/Experimental source: strain PA01  
C/Genetics:  
A/Gene: PA2306

Query Match	27.4%;	Score 272.5;	DB 2;	Length 205;
Best Local Similarity	37.0%;	Pred. No. 3.6e-17;		
Matches	71;	Conservative 39;	Mismatches 67;	Indels 15; Gaps 8;

  

QY	5	LLSAFWTYTLITAMTPGPNNTIIALSSATSHGFRQSTRVLGMSLGLIYMLLCAGISFSL	64
		: : :       : : :   : :   : :    : : :	
Db	7	LLAFLSLFAFVTSVTPGPNNTMLLASGVNFGFVRSIPIHLIGISCGFIMVM--AVGFGL	63
		: : :       : : :   : :   : :    : : :	
QY	65	AVI--DPAAVHLISWAGAAIYIVWLAKIATS-PTKEDGL-QAKPISFWASFALQFVNVK	119
		: : :   : :       : :         : :   : :   : :	
Db	64	GTVEEAYPVLYTLIRYGAAYLLYLAWKIATSGPAGNDPEGRGKPLSYWGAAPQWVNP	123
		: : :   : :       : :         : :   : :   : :	

```

OY      120 ILLGVTAALSTFVLPTQALSWMVVGVSULLAMIG--TFGNVCWALAGHLFORLFR--QYG 175
           : |||::| | : || : || : || | | | | : : : : : :
Db      124 AWMVAWGAIISTYT-PLGGYFTNVVVISTVFALINAPTIG--IWAGFGSMLRNVRADPRWL 180

OY      176 RQLNIVLALLLV 187
           | | : |||||
Db      181 RVENVGMALLLV 192

RESULT 6
A/I0022
Probable ABC-transport protein YP00181 [imported] - Yersinia pestis (strain CO92)
C/Species: Yersinia pestis
C/Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Jul-2004
C/Acession: AI0022
R/Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall,
Nature 413, 523-527, 2001
A/Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A/Reference number: AB0001; MUID:21470413; PMID:11586360
A/Acession: AI0022
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-200 <KUR>
A/Cross-references: UNIPROT:Q8ZJD2; GB:AL590842; PIDN:CAC89043.1; PID:g15978284; GSPDB:G
C/Genetics:
A/Gene: YP00181

Query Match          23.6%; Score 235.5; DB 2; Length 200;
Best Local Similarity 30.5%; Pred. No. 6.9e-14;
Matches 60; Conservative 44; Mismatches 84; Indels 9; Gaps 3;

OY      4 TLLSAFWTYTLITAMTPGNNILALSSATSHGFROSTRVLAGMSLGFLIYMLCAGISFS 63
           ::::| : |||||::| : : || : : || : : || : : ||
Db      8 SWVGFLW---VAAITPGNNMLLTSTGANFGFMRSIMLMGLMGQSMILLVAFGVGS 63

OY      64 LAVIDPAAVHLLSWAGAAYIWLAWKIATSPTKE--DGLQAKPISEFWASFAIQFNVKI 120
           | : | : | : | : | : | : | : | : | : | : | : | : |
Db      64 LLIYPSLHALKILSGSYLLMWLWKIATSAYEKLDITNAVAPAKPLRLYQGWLQFLNPKA 123

OY      121 ILGYVTALSTFVLPTQALSWMVVGVSULLAMIGTFGNVCWALAGHLFORLF--ROYGRQL 178
           | | : : : : | : : : : : : : : : : | | | | : :
Db      124 WLWGLGSVASFSMAGAAYNHSILMISLGMFMVNVLVAGIIWLGFGLIGRLLSRRAWFTF 183

OY      179 NIVLALLLVCAVRIFY 195
           || : ||| | : || :
Db      184 NISMGLLTAACTVLLIWH 200

RESULT 7
B82166
conserved hypothetical protein VC1712 [imported] - Vibrio cholerae (strain N16961 serogru
C/Species: Vibrio cholerae
C/Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004
C/Acession: B82166
R/Heldelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, F
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A/Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A/Reference number: AB2035; MUID:20406833; PMID:10952301
A/Acession: B82166
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-204 <HEI>
A/Cross-references: UNIPROT:Q9KRD0; GB:AE004249; GB:AE003852; NID:g9656228; PIDN:AAF9486
C/Genetics:
A/Gene: VC1712
A/Map position: 1

```

```

Query Match      17.3%; Score 172; DB 2; Length 204;
Best Local Similarity 29.9%; Pred. No. 3.3e-08;
Matches 49; Conservative 36; Mismatches 71; Indels 8; Gaps 5;

QY      4 TLLSAFWYTYLTITAMTPGPNNTILSSATSHGFRQSTRVLAGMSLGFLLVMLLCAGISFS 63
      || : : : ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      14 TLFGLPLAFAFVSTFSPGPNNTIMMTSGAVNGFLRTIPHMLGITFGFSINVLV-VGVGLT 72

QY      64 -LAVIDPAAVHLLSWAGAAIYVWLAMKIATSPTEKEDGLQAKPISFWASPALQFVNKIIIL 122
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      73 ELFORYPILQGGIQLICTLLVLYLVAVKIALSRPSKEGEYQPMSFVAALMFGWVNPCKGWS 132

QY      123 XGVTALSTFVLPQTQALSWV-VG-VSVLLAMIGTFGNVCWALAG 164
      : || : | | : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      133 MALTAVSVF----NENASWLQGLIALVFALVNLPSVSATWTAAG 172

RESULT 8
E87264
efflux protein, lyse family [imported] - Caulobacter crescentus
C/Species: Caulobacter crescentus
C/Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Jul-2004
C/Accession: E87264
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolo-
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A/Title: Complete Genome Sequence of Caulobacter crescentus.
A/Reference number: A87249; MUID:21173698; PMID:11259647
A/Accession: E87264
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-222 <STO>
A/Cross-references: UNIPROT:Q9ABU4; GB:AB005673; NID:g13421235; PIDN:AAK22113.1; GSPDB:
C/Genetics:
A/Gene: CC0126

Query Match      17.0%; Score 169; DB 2; Length 222;
Best Local Similarity 28.6%; Pred. No. 6.5e-08;
Matches 58; Conservative 34; Mismatches 93; Indels 18; Gaps 5;

QY      1 VTPTLSAFWYTYLTITAMTPGPNNTILSSATSHGFRQSTRVLAGMSLGFLLVMLLCAGI 60
      || | | | | | : : ||||| : : : : : : : : : : : : : : : : : : : : : :
DB      18 VDPARYGAFLVAMFVMAITPGPANLFAIATGMEKGKGAALAGVAGMNCATLIWIFIGSALG 77

QY      61 SFSLAVIDPAAVHLLSWAGAAIYVWL-----AWKIATSPTEKEDGLQAKPISFWASPAL 113
      : : : : : | | | | | : : | | | | | | | | | | | | | | | | | | | | | |
DB      78 LGLALIAFPQAFHLALAGAAIYVWLGLKSLWAGWKNASHAVA-AVRGKSAFVDGFAV 136

QY      114 QFVNVKIILYGTALSTF-----VLPTQALSWVGVSVLLAMIGTFGNVCWALAGHLF 167
      | | | | | : : : : : | | | | | : : : : : : : : : : : : : : : : : : :
DB      137 QIANPKALLFTAVLPPFLVDVDRPLVPQLMFACAT-IGMDLISMSAVGVGAALSDRMN 195

QY      168 QRLFRQYGRQLNTVLLALLVYCA 190
      : | | | | | : : : : : | | : : : : : : : : : : : : : : : : : : :
DB      196 EPGFR---RGFAILVGVLLITAA 215

RESULT 9
B91223
homoserine/homoserine lactone efflux protein [imported] - Escherichia coli (strain O15
C/Species: Escherichia coli
C/Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C/Accession: B91223
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A/Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen
A/Reference number: A99629; MUID:21156231; PMID:11258796
A/Accession: B91223
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-206 <HAY>

```



A;Cross-references: UNIPROT:P27847; GB:BA000007; PIDN:BAB38177.1; PID:g13364230; GSPDB:G  
A;Experimental source: strain O157:H7, substrain R1MD 0509952  
C;Genetics:  
A;Gene: ECs4754

Query Match	16.6%;	Score 165.5;	DB 2;	Length 206;
Best Local Similarity	27.0%;	Pred. No. 1.2e-07;		
Matches	53;	Conservative	42;	Mismatches 80; Indels 21; Gaps 9;
QY	8	AFWTYTLITAMTPGPNNILALSSATSHGFQRSTRYLAGMSLGFLIYMLCA---GISFSL	64	
Db	8	AYLTSIIILSPGSGAINTMTTSLNHGYRGAVASIAQLQTGLAIHIVLVGVGLGTLFSR	67	
QY	65	AVIDPAVHLISWAGAYIVWLAA--WKIATSPFKED--GIQAKPISFWASFALQFVNVK	119	
Db	68	SVI---AFEVLKMWAGAYLWIWLGIQOWRAAGALDKLASTQSRRHLFORAVEVNLTNPK	124	
QY	120	IILYGVTALSTFVL-P-QTQALSWMV-GVSV----LLAMIG--TFGN--VCNALAGHLFOR	169	
Db	125	SIVFLAALFPQFIMPOQPOLMQYIVLVGVTTVIVDIWMIGYATLAQRIALMWPCKOMKA	184	
QY	170	LFRÖYGRÖLNIVLALL	185	
Db	185	INKIFGSLFMVLVGALL	200	

RESULT 10  
AI2614  
RhTB family transporter rhTB [imported] - Agrobacterium tumefaciens (strain C58, Dupont)  
C/Species: Agrobacterium tumefaciens  
C/Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 09-Jul-2004  
C/Accession: AI2614  
R/wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.  
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kuttyavin, T.; Levy, R.; Li, M.; Mclellan,  
; Kaup, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,  
ster, E.W.  
A/Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A/Reference number: AB2577; MUID:21608550; PMID:11743193  
A/Accession: AI2614  
A/Status: Preliminary  
A/Molecule type: DNA  
A/Residues: 1-205 <KUR>  
A/Cross-references: UNIPROT:Q8U1I2; GB:AE008688; PIDN:AAL41335.1; PID:gl7738647; GSPDB:G  
A/Experimental source: strain C58 (Dupont)  
C/Genetics:  
A/Gene: rhTB  
A/Map position: circular chromosome

Query Match 16.5%; Score 164; DB 2; Length 205;  
Best Local Similarity 29.6%; Pred. No. 1.7e-07;  
Matches 50; Conservative 33; Mismatches 66; Indels 20; Gaps 4;

```
QY      1 VTPTLLSAFWTTYTLITAMTPGPNNILASSATSHGFROSTRVLAGMSLGLIYVMLLCAGI 60  
        :| | | |: | | | | | | :: | | | | : | | | | :  
Db      1 MTLTLLAYATAALFIAAIPGPMTAIVARALGSGFRFTFFMGLGLVLGDMYL---TGV 57  
  
QY      61 SFSLAVIDPA---AVHLLSMAGAAYIVWLAWKIATSPTRKEDGLQAK-----PISFMASFA 112  
        || : | : : | | | | : : | | : | : | : | : | :  
Db      58 ILGLAFVAQTFOEAFMWLVKFKAGAAYLIIYIAWKLMWTAGLLPQDLKARKSTSIPIWFLSGLL 117  
  
QY      113 LQFNVKIILGYVTALSTFY-----LPQTQALSWMVGVSVLLAMI 152  
        : | | : | : | : | : | | | | | | | | | | | |  
Db      118 ITLGNPKTMLEFVALVPRTLIDIRIMIGPSEYATLLALTFFVLMVALLPYI 166
```

RESULT 11  
H97396  
dihydrodipicolinate reductase (AP001519) [imported] - Agrobacterium tumefaciens (strain  
C;Species: Agrobacterium tumefaciens  
C;Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 09-Jul-2004  
C;Accession: H97396

R; Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Iappas, C.; Markelz, B., Science 294, 2323-2328, 2001  
A; Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent *Agrobacterium tumefaciens* A; Reference number: A97359; MUID:21608551; PMID:11743194

A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-224 <KUR>  
A;Cross-references: UNIPROT:Q8U1I2; GB:AE007869; PIDN:AAK86129.1; PID:g15155214; GSPDB:  
C;Genetics:  
A;Gene: AGR C 546  
A;Map position: circular chromosome

[illegible]

RESULT 12  
E87252  
efflux protein, Lyse family [imported] - Caulobacter crescentus  
C;Species: Caulobacter crescentus  
C;Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 09-Jul-2004  
C;Accession: E87252  
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.  
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon  
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A;Title: Complete Genome Sequence of Caulobacter crescentus.  
A;Reference number: A87249; MUID:21173698; PMID:11259647  
A;Accession: E87252  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-210 <STO>  
A;Cross-references: UNIPROT:Q9AC37; GB:AE005673; NID:g13421119; PIDN:AAK22017.1; GSPDB:((  
C;Genetics:  
A;Gene: CC0029  
C;Superfamily: hypothetical protein b1798

	Query Match	16.4%;	Score 163;	DB 2;	Length 210;
	Best Local Similarity	29.2%;	Pred. No. 2.1e-07;		
	Matches	Conservative	29;	Mismatches	77;
				Indels	20;
				Gaps	7;
QY	8 AFWTYTLITAMTPGPNNTIALSSATSHGFROSTRVLAGMSLGLIIVMLCAGISF-SLAV 66     :     : :   :   :   :   :   :				
Dd	9 AFLVLCSMALTGGNNMVLVRSICQGRWAGIVSLIGTAAGF-VVYLICALGITALLM 67				
QY	67 IDPAVHLLSWAGAYIVWLAMKIATSPTKEDGLQAKPIS-----FWASFAQLQFVNVK 119       :   :   :   :   :   :   :   :				
Dd	68 AAPAIAYDLIRFGCALYLAWLAWQ-AIRPGASPFQVRRLPKDSPAKLTIMGEVTNMLNPK 126				
QY	120 IILYGVALSTFVLEPQ-----TQALSWVGVSVILLAMIGTFGNVCWA---LAGHLFOR 169 : : : :   : : :   : : : : :   :     :				
Dd	127 AAMLVLSLLPQFIKEHGCVFTOSLA--LGLTQICVSLTVNGVICLAAGTIAAGFLAAR 182				

RESULT 13  
AG3485  
homoserine/homoserine lactone efflux protein [imported] - Brucella melitensis (strain 1  
C;Species: Brucella melitensis

C;/Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 09-Jul-2004  
C;/Accession: AG3485  
R;/DelVecchio, V.G.; Kapatal, V.; Redkar, R.J.; Patra, G.; Mujeer, C.; Ios, T.; Ivanova,  
.; Mazur, M.; Goldsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess  
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002  
A;/Title: The genome sequence of the facultative intracellular pathogen *Brucella melitensis*  
A;/Reference number: AD3252; PMID:11756688  
A;/Accession: AG3485  
A;/Status: preliminary  
A;/Molecule type: DNA  
A;/Residues: 1-249 <KUR>  
A;/Cross-references: UNIPROT:Q8YEK9; GB:AE008917; PIDN:AAL53050.1; PID:g17983910; GSPDB:G  
A;/Experimental source: strain 16M  
C;/Genetics:  
A;/Gene: BMEI1869  
A;/Map position: I

Query Match	16.3%	Score 162.5;	DB 2;	Length 249;
Best Local Similarity	32.3%	Pred. No. 2.8e-07;		
Matches 50; Conservative	22;	Mismatches 72;	Indels 11;	Gaps 5;

Qy 4 TLLSAFWTTYTLITAMTPGPNNILALSSATSHGFRQSTRVLGMSLGLI-VMLLCAGISF 62

Db 44 TIFVQFAIATAILSITPGPDMTLFVGRALSEKKAAGFACMAGASTGIIVHTSMVALGLS- 102

```

Oy      63 SLAVIDPAVHLLSWAGAAIYIWLAWKI-----ATSPTEKEDGLQAKPISFW-ASFALQFV 116
        :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      103 ALIVASPAFTALKVVGAGYLVWLAVQAI CKGSAFSPKXNGGKHTLFQNNLTGLGILNL 162

```

```
QY 117 NKIIILGVLTALSTFVL--PQTALSMWVGVSVL 148
      | | | | : : | | | : : | :
Db 163 NPKIILENMTFLPQFVSADHDPHMGKLPFLGLSFI 197
```

RESULT 14  
F82353  
conserved hypothetical protein VC0191 [imported] - *Vibrio cholerae* (strain N16961 serogroup O1)  
C/Species: *Vibrio cholerae*  
C/Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 09-Jul-2004  
C/Accession: F82353  
R;Heldelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.; Chardonson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, F.; R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
Nature 406, 477-483, 2000  
A;Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.  
A;Reference number: AB2035; MUID:20406833; PMID:10952301

A/Accession: F82353  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-222 <HEI>  
A/Cross-references: UNIPROT:Q9KVF5; GB:AE004109; GB:AE003852; NID:g9654590; PIDN:AAF9336  
A/Experimental source: serogroup O1; strain NI6961; biotype EI Tor  
C/Genetics:  
A/Gene: VC0191  
A/Map position: 1  
C/Superfamily: hypothetical protein b1798

Query Match	16.1%;	Score 160.5;	DB 2;	Length 222;
Best Local Similarity	25.0%;	Pred. No. 3.8e-07;		
Matches 47;	Conservative 45;	Mismatches 79;	Indels 17;	Gaps 6;

Oy 15 ITAMTPGPNNIILASSATSHGFROSTRVLAGMSLGLIVMLL-CAGISFSLAVIDPAVH 73  
|::||:|::|||::|::||:|::||:|::||:|:  
Db 16 IALLSPGPDVALVQONATQHGRKTGMIALGLSCGILVHLILSLSGISY-LVKOOPMLEN 74

```

Oy      74 LLSWGAAYIVWL-----AMKIATSPTEKDG--LQAKPISFWASFALQFVNVKII 121
          ||  ||  ::::|  ||  ||  ::::|  ||  ||  ::::|  ||  ||  ::::|
Db      75 LLQLAGSYLLYLQAGALQSVMAQKNASTPTHSAPSIIGNRRQAFYTKGMNTLLNLPKAL 134

```

Oy      122 LYGVTALSTFVLPTQTALSMWVGSVLLAMIGT--FGNVCMALAGHLFORLEFRQYGRQLN 179  
::|::||:::||::||  
  
Db    135 VFVSLSS-LIPASMSVSGKSAAILVGLSITWFSCLAWLTTSAMQQRMQRITRSVD 193

QY	180	IVLALLV	187
	:		:
Db	194	SICAVFI	201

RESULT 15  
B83280

hypothetical protein PA2916 [imported] - *Pseudomonas aeruginosa* (strain PAO1)  
C;Species: *Pseudomonas aeruginosa*

C;Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 09-Jul-2004  
C;Accession: B83280

R.; Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Badman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim

..; Lory, S.; Olbon, M.V.  
Nature 406, 959-964, 2000

A;Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen  
A;Reference number: AB2950; MUID:20437337; PMID:10984043

A/Accession: B83280  
A/Status: preliminary

A;Molecule\_type: DNA  
A;Residues: 1-197 <STO>

A; Cross-references: UNIPROT:Q9H2T1, GB:AE004718, GB:AE004091, NID:g99949009, PIDN:AAG06301  
A; Experimental source: strain PA01

C;Genetics:  
A;Gene: PA2916

Query Match	15.9%;	Score 158.5;	DB 2;	Length 197;
Best Local Similarity	27.0%;	Pred. No. 5e-07;		
Matches	51;	Conservative	35;	Mismatches 98;
				Indels 5;
				Gaps 4;

Qy 6 LSAFWITYLLTAMTPGPNNILALSSATSHGFRQSTRVLAGMSLGF-LIWMLICAGISFSL 64  
| : : : | | | | : : : | : : : |  
Db 4 LVPFLLFAVVASITPGFTNIIVLNSQRHGLAAMPVLVGACAAVALILLGLG-EL 62

Qy 65 AVIDPAVHLTSMGAAYIVLAWKIATSPTEKEDLQ-AKPIISFWASFALQEVNVKILY 1233

Db 63 LRRHPLQOGLAWLGVGHTSYLAWSLFRSAGCIDGAEPFRRLGVLTGGAALQLVNPKAMM 1222

```

Qy      124  GVTALSTFVLPTQTALSWVGVSVLLAMIGTFGNVCWALAGHLFQRLFRQY--RQLNIV 1811
      : ||: |      : ::| : : ||| | ||| | : : |
Db      123  ALAALAFAGEGAGQAGRIGILLALLFFLVSLPCLASWALLGVGSARLLPSLMKRFNQ 1822

```

QY	182 LALLVYCA 190
	:   :
Db	183 MALLLASA 191

Search completed: September 10, 2005, 01:45:55  
Job time : 42 secs

***This Page Blank (uspto)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2005, 01:34:51 ; Search time 173 Seconds  
(without alignments)  
577.199 Million cell updates/sec

Title: US-10-620-487-2

Perfect score: 996

Sequence: 1 VTPTLSAFWYTLITAMP.....RQLNIVLALLVCAVRIFY 195

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : UniProt\_03:\*  
1: uniprot\_sprot:\*  
2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	993	99.7	195	1 YFIK_ECOLI	P38101 escherichia
2	989	99.3	195	2 Q8FFI1	Q8ffl1 escherichia
3	987	99.1	195	2 Q83K22	Q83k22 shigella fl
4	979	98.3	195	2 Q8XA19	Q8xa19 escherichia
5	884	88.8	195	2 Q8Z4J7	Q8z4j7 salmonella
6	884	88.8	195	2 Q8ZMX5	Q8zmx5 salmonella
7	367.5	36.9	200	2 Q7P2H4	Q7p2h4 fusobacteri
8	316.5	31.8	194	2 Q63BD0	Q63bd0 bacillus ce
9	311.5	31.3	194	2 Q81QH8	Q81qh8 bacillus an
10	311.5	31.3	194	2 Q6HIR1	Q6hir1 bacillus th
11	310.5	31.2	194	2 Q738B5	Q738b5 bacillus ce
12	303.5	30.5	194	2 Q81DI7	Q81di7 bacillus ce
13	293	29.4	192	2 Q65N28	Q65n28 bacillus li
14	277.5	27.9	198	2 Q82971	Q82971 bacillus sp
15	272.5	27.4	205	2 Q911G9	Q911g9 pseudomonas
16	262.5	26.4	209	2 Q88HF4	Q88hf4 pseudomonas
17	258.5	26.0	202	2 Q87ZH7	Q87zh7 pseudomonas
18	256	25.7	202	2 Q7N9E0	Q7n9e0 photorhabdu
19	244	24.5	204	2 Q6LKF4	Q6lkf4 photobacter
20	237.5	23.8	200	2 Q6CZU2	Q6czu2 erwinia car
21	236	23.7	219	2 Q7W450	Q7w450 bordetella
22	236	23.7	219	2 Q7WFK7	Q7wfk7 bordetella
23	235.5	23.6	197	2 Q92NS0	Q92ns0 rhizobium m
24	235.5	23.6	200	2 Q664P6	Q664p6 yersinia ps
25	235.5	23.6	200	2 Q8ZJD2	Q8zjd2 yersinia pe
26	234.5	23.5	192	2 Q8CK91	Q8ck91 yersinia pe
27	232.5	23.3	202	2 Q89G49	Q89g49 bradyrhizob
28	232	23.3	219	2 Q7VSL5	Q7vsl5 bordetella
29	231	23.2	196	2 Q7QMF4	Q7qmf4 anopheles g
30	222.5	22.3	197	2 Q8EG90	Q8eg90 shewanella
31	221	22.2	193	2 Q88HC2	Q88hc2 pseudomonas

32	215.5	21.6	199	2 Q8D574	Q8d574 vibrio vuln
33	214.5	21.5	188	2 Q882W8	Q882w8 pseudomonas
34	213.5	21.4	198	2 Q98KK8	Q98kk8 rhizobium 1
35	211.5	21.2	200	2 Q6LVD7	Q6lvd7 photobacter
36	211.5	21.2	206	2 Q7NWS8	Q7nws8 chromobacte
37	206.5	20.7	208	2 Q7NWP5	Q7nwp5 chromobacte
38	206	20.7	195	2 Q7MG87	Q7mg87 vibrio vuln
39	204.5	20.5	248	2 Q7WDP6	Q7wdp6 bordetella
40	202	20.3	195	2 Q8D4N7	Q8d4n7 vibrio vuln
41	202	20.3	205	2 Q62EH2	Q62eh2 burkholderi
42	202	20.3	205	2 Q63P95	Q63p95 burkholderi
43	199.5	20.0	208	2 Q7W2P8	Q7w2p8 bordetella
44	185.5	18.6	209	2 Q7VS75	Q7vs75 bordetella
45	183	18.4	194	2 Q87K08	Q87k08 vibrio para

ALIGNMENTS

RESULT 1  
YFIK\_ECOLI  
ID YFIK\_ECOLI STANDARD; PRT; 195 AA.  
AC P38101;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 25-JAN-2005 (Rel. 46, Last annotation update)  
DE Hypothetical protein yfik.  
GN Name=yfik; OrderedLocustNames=b2578;  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RA Nashimoto H.;  
RT "Non-ribosomal proteins affecting the assembly of ribosomes in  
RT Escherichia coli.";  
RL (In) Nierhaus K.H. (eds.);  
RL The translational apparatus, pp.185-195, Plenum Press, New York  
RL (1993).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RA Nashimoto H.; Saito N.;  
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503; DOI=10.1126/science.277.5331.1453;  
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12";  
RL Science 277:1453-1474(1997).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=97349980; PubMed=9205837;  
RA Yamamoto Y., Aiba H., Baba T., Hayaishi K., Inada T., Isono K.,  
RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,  
RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,  
RA Oshima T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasubram S.,  
RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,  
RA Yamagata S., Horiuchi T.;  
RT "Construction of a contiguous 874-kb sequence of the Escherichia coli-  
RT K12 genome corresponding to 50.0-68.8 min on the linkage map and  
RT analysis of its sequence features";  
RL DNA Res. 4:91-113(1997).  
RN [5]  
RP IDENTIFICATION.  
RA Rudd K.E.;



RL Unpublished observations (AUG-1994).  
CC -I- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
CC -I- SIMILARITY: Belongs to the rht family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; D13169; -; NOT\_ANNOTATED\_CDS.  
DR EMBL; D64044; -; NOT\_ANNOTATED\_CDS.  
DR EMBL; U00096; AAC75631.1; -.  
DR EMBL; D90886; BAA16464.1; -.  
DR PIR; A65036; A65036.  
DR Echobase; EB2339; -.  
DR EcoGene; EG12445; yfik.  
DR InterPro; IPR001123; Lyse.  
DR Pfam; PF01810; Lyse; 1.  
KM Complete proteome; Hypothetical protein; Transmembrane.  
FT TRANSMEM 8 28 Potential.  
FT TRANSMEM 47 67 Potential.  
FT TRANSMEM 71 91 Potential.  
FT TRANSMEM 142 162 Potential.  
SQ SEQUENCE 195 AA; 21248 MW; 5F86B828DDDEC090 CRC64;

Query Match 99.7%; Score 993; DB 1; Length 195;  
Best Local Similarity 99.5%; Pred. No. 5.8e-73;  
Matches 194; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VTPTLLSAFWTYTLITAMTPGPNNILALSSATSHGFRQSTRVLAGMSLGFLIVMLCAGI 60  
:|||||  
Db 1 MTPTLLSAFWTYTLITAMTPGPNNILALSSATSHGFRQSTRVLAGMSLGFLIVMLCAGI 60  
  
QY 61 SFSLAVIDPAVHLLSWAGAYIVWLAWKIATSPTKEDGLQAKPISFWASPALQFVNVKI 120  
|||||  
Db 61 SFSLAVIDPAVHLLSWAGAYIVWLAWKIATSPTKEDGLQAKPISFWASPALQFVNVKI 120  
  
QY 121 ILGYVTALSTFVLPTQALSWVGVSVLLAMIGTFGNVCWALAGHLFQRLFRQYGRQLNI 180  
|||||  
Db 121 ILGYVTALSTFVLPTQALSWVGVSVLLAMIGTFGNVCWALAGHLFQRLFRQYGRQLNI 180  
  
QY 181 VLALLLVYCAVRIFY 195  
|||||  
Db 181 VLALLLVYCAVRIFY 195

RESULT 2  
Q8FF11 PRELIMINARY; PRT; 195 AA.  
ID Q8FF11  
AC Q8FF11; 01-MAR-2003 (Tremblrel. 23, Created)  
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)  
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
DE Hypothetical protein yfik.  
GN Name=yfik; OrderedLocustNames=c3102;  
OS Escherichia coli O6.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=217992;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=O6:H1 / CFT073 / ATCC 700928 / UPEC;  
RX MEDLINE=22388234; PubMed=12471157; DOI=10.1073/pnas.252529799;  
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,  
RA Raske D., Buckles E.L., Lion S.-R., Boutin A., Hackett J., Stroud D.,  
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,  
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;  
RT "Extensive mosaic structure revealed by the complete genome sequence  
RT of uropathogenic Escherichia coli.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).

DR EMBL; AE016764; AAN81551.1; -.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0005293; F:lysine permease activity; IEA.  
DR GO; GO:0006865; P:amino acid transport; IEA.  
DR InterPro; IPR001123; Lyse.  
DR Pfam; PF01810; Lyse; 1.  
KM Complete proteome; Hypothetical protein.  
SQ SEQUENCE 195 AA; 21278 MW; 295DD583ADCA8584 CRC64;

Query Match 99.3%; Score 989; DB 2; Length 195;  
Best Local Similarity 99.0%; Pred. No. 1.2e-72;  
Matches 193; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VTPTLLSAFWTYTLITAMTPGPNNILALSSATSHGFRQSTRVLAGMSLGFLIVMLCAGI 60  
:|||||  
Db 1 MTPTLLSAFWTYTLITAMTPGPNNILALSSATSHGFRQSTRVLAGMSLGFLIVMLCAGI 60  
  
QY 61 SFSLAVIDPAVHLLSWAGAYIVWLAWKIATSPTKEDGLQAKPISFWASPALQFVNVKI 120  
|||||  
Db 61 SFSLAVIDPAVHLLSWAGAYIVWLAWKIATSPTKEDGLQAKPISFWASPALQFVNVKI 120  
  
QY 121 ILGYVTALSTFVLPTQALSWVGVSVLLAMIGTFGNVCWALAGHLFQRLFRQYGRQLNI 180  
|||||  
Db 121 ILGYVTALSTFVLPTQALSWVGVSVLLAMIGTFGNVCWALAGHLFQRLFRQYGRQLNI 180  
  
QY 181 VLALLLVYCAVRIFY 195  
|||||  
Db 181 VLALLLVYCAVRIFY 195

RESULT 3  
Q83K22 PRELIMINARY; PRT; 195 AA.  
ID Q83K22  
AC Q83K22; Q7C0E0;  
DT 01-JUN-2003 (Tremblrel. 24, Created)  
DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)  
DT 25-OCT-2004 (Tremblrel. 28, Last annotation update)  
DE Hypothetical protein yfik.  
GN Name=yfik; OrderedLocustNames=S2813, SF2640;  
OS Shigella flexneri.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Shigella.  
OX NCBI\_TaxID=623;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=301 / Serotype 2a;  
RX MEDLINE=22272406; PubMed=12384590; DOI=10.1093/nar/gkf566;  
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,  
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,  
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,  
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,  
RA Yu J.;  
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity  
RT through comparison with genomes of Escherichia coli K12 and O157.";  
RL Nucleic Acids Res. 30:4432-4441(2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=2457T;  
RX MEDLINE=22590274; PubMed=12704152;  
RX DOI=10.1128/IAI.71.5.2775-2786.2003;  
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,  
RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,  
RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,  
RA Schwartz D.C., Blattner F.R.;  
RT "Complete genome sequence and comparative genomics of Shigella  
RT flexneri serotype 2a strain 2457T.";  
RL Infect. Immun. 71:2775-2786(2003).  
DR EMBL; AE015279; AAN44136.1; -.  
DR EMBL; AE016987; AAP17960.1; -.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0005293; F:lysine permease activity; IEA.  
DR GO; GO:0006865; P:amino acid transport; IEA.  
DR InterPro; IPR001123; Lyse.

DR Pfam; PF01810; Lyse; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 195 AA; 21260 MW; D77D7B0A152ED16C CRC64;

Query Match 99.1%; Score 987; DB 2; Length 195;  
Best Local Similarity 99.0%; Pred. No. 1.8e-72;  
Matches 193; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VTPTLSAFWYTYLTITAMTPGPNNILALSSATSHGFRQSTRVLGMSLGFLIVMLCAGI 60  
Db 1 MTPILLSAFWYTYLTITAMTPGPNNILALSSATSHGFRQSTRVLGMSLGFLIVMLCAGI 60

QY 61 SFSLAVIDPAAVHLISWAGAAIYVWLAKIATSPTEKEDGLQAKPISFWASFALQFVNVKI 120  
Db 61 SFSLAVIDPAAVHLISWAGAAIYVWLAKIATSPTEKEDGLQAKPISFWASFALQFVNVKI 120

QY 121 ILVGTALSTFVLPTQALSWVGVSVLLAMIGTFGNVCWALAGHLFQRLFRQYGRQLNI 180  
Db 121 ILVGTALSTFVLPTQALSWVGVSVLLAMIGTFGNVCWALAGHLFQRLFRQYGRQLNI 180

QY 181 VLALLLVYCAVRIFY 195  
Db 181 VLALLLVYCAVRIFY 195

RESULT 4  
Q8XA19 PRELIMINARY; PRT; 195 AA.

AC Q8XA19; Q7ABJ3;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)  
DE Hypothetical protein yfik (Hypothetical protein ECs3444).  
GN Name=yfik; OrderedlocusNames=ECs3444, z3861;  
OS Escherichia coli O157:H7.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=83334;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=O157:H7 / BDL933 / ATCC 700927 / EHEC;  
RX MEDLINE=21074935; PubMed=11206551; DOI=10.1038/35054089;  
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouis K.,  
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
RA Welch R.A., Blattner F.R.;  
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";  
RL Nature 409:529-533 (2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=O157:H7 / RIMD 0509952 / EHEC;  
RX MEDLINE=21156231; PubMed=11258796;  
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;  
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli  
O157:H7 and genomic comparison with a laboratory strain K-12.";  
RL DNA Res. 8:11-22 (2001).  
DR EMBL; AE005488; AAG57694.1; -.  
DR EMBL; AP002562; BAB36867.1; -.  
DR PIR; B85904; B85904.  
DR PIR; D91059; D91059.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0005293; F:lysine permease activity; IEA.  
DR GO; GO:0006865; P:amino acid transport; IEA.  
DR InterPro; IPR001123; Lyse.  
DR Pfam; PF01810; Lyse; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 195 AA; 21301 MW; 54DA47EF23488F07 CRC64;

Query Match 98.3%; Score 979; DB 2; Length 195;  
Best Local Similarity 96.9%; Pred. No. 7.9e-72;  
Matches 189; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 VTPTLSAFWYTYLTITAMTPGPNNILALSSATSHGFRQSTRVLGMSLGFLIVMLCAGI 60  
Db 1 MTPILLSAFWYTYLTITAMTPGPNNILALSSATSHGFRQSTRVLGMSLGFLIVMLCAGI 60

QY 61 SFSLAVIDPAAVHLISWAGAAIYVWLAKIATSPTEKEDGLQAKPISFWASFALQFVNVKI 120  
Db 61 SFSLAVIDPAAVHLISWAGAAIYVWLAKIATSPTEKEDGLQAKPISFWASFALQFVNVKI 120

QY 121 ILVGTALSTFVLPTQALSWVGVSVLLAMIGTFGNVCWALAGHLFQRLFRQYGRQLNI 180  
Db 121 ILVGTALSTFVLPTQALSWVGVSVLLAMIGTFGNVCWALAGHLFQRLFRQYGRQLNI 180

QY 181 VLALLLVYCAVRIFY 195  
Db 181 VLALLLVYCAVRIFY 195

RESULT 5  
Q8Z4J7 PRELIMINARY; PRT; 195 AA.

AC Q8Z4J7; Q7CBN2;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)  
DE Putative membrane protein.  
GN Name=yfik; OrderedlocusNames=STY2838, t0265;  
OS Salmonella typhi.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Salmonella.  
OX NCBI\_TaxID=601;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CT18;  
RX MEDLINE=21534947; PubMed=11677608; DOI=10.1038/35101607;  
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,  
RA Churcher C.M., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,  
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,  
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,  
RA Felwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,  
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,  
RA Quail M.A., Rutherford K.M., Simmonds M., Skelton J., Stevens K.,  
RA Whitehead S., Barrett B.G.;  
RT "Complete genome sequence of a multiple drug resistant Salmonella  
enterica serovar Typhi CT18.";  
RL Nature 413:848-852 (2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Ty2 / ATCC 700931;  
RX MEDLINE=22531367; PubMed=12644504;  
DOI=10.1128/JB.185.7.2330-2337.2003;  
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,  
RA Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;  
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2  
and CT18.";  
RL J. Bacteriol. 185:2330-2337 (2003).  
DR EMBL; AL627275; CAD02794.1; -.  
DR EMBL; AE016835; AAO67991.1; -.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0005293; F:lysine permease activity; IEA.  
DR GO; GO:0006865; P:amino acid transport; IEA.  
DR Pfam; PF01810; Lyse; 1.  
KW Complete proteome.  
SQ SEQUENCE 195 AA; 21334 MW; EF30142B85995E85 CRC64;

Query Match 88.8%; Score 884; DB 2; Length 195;  
Best Local Similarity 83.6%; Pred. No. 4.1e-64;  
Matches 163; Conservative 19; Mismatches 13; Indels 0; Gaps 0;

QY 1 VTPTLSAFWYTYLTITAMTPGPNNILALSSATSHGFRQSTRVLGMSLGFLIVMLCAGI 60

Db 1 MTPMLLSAFWYTYLTALTTPGPNNILALSAATAHGFRQSIKRVLAGMSLGFLLVMLLCAGI 60  
QY 61 SFSLAVIDPAVAHLLSWAGAAYIVWLAWKIATSPTEKEDGLQAKPISFWASFALQFVN VKI 120  
Db 61 AFSLAVIDPAIITHLSWGAAYIILWLAWKIATSPADEKVRPKPVGFVWSFGLQFVN VKI 120  
QY 121 ILXGVTALSTFVLPQTOALSWVGVSVLLAMIGTFGNVCWALAGHLFQRLFRQYGRQ LNI 180  
Db 121 ILXGITALSTFVLPQTOALNWVIGVSILLALIGTFGNVCWALAGHLFQRAFRHYGRQ LNI 180  
QY 181 VIALLLVYCAVRIFY 195  
Db 181 ILALLLVYCAVRIFY 195

RESULT 6  
Q8ZMK5 PRELIMINARY; PRT; 195 AA.  
ID Q8ZMK5  
AC Q8ZMK5;  
DT 01-MAR-2002 (Tremblrel. 20, Created)  
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)  
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
DE Putative transport protein.  
GN Name=yfik; OrderedLocusNames=STM2645;  
OS Salmonella typhimurium.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Salmonella.  
OX NCBI\_TaxID=602;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=LT2;  
RX MEDLINE=21534948; PubMed=11677609; DOI=10.1038/35101614;  
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
RA Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,  
RA Waterston R., Wilson R.K.;  
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium  
RT LT2.";  
RT Nature 413:852-856 (2001).  
RL EMBL; AE008820; AAL21539.1; -  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0005293; F:lysine permease activity; IEA.  
DR GO; GO:0006865; P:amino acid transport; IEA.  
DR Pfam; PF01810; Lyse; 1.  
KW Complete proteome.  
SQ SEQUENCE 195 AA; 21292 MW; B63E83C2FB104EE3 CRC64;

Query Match 88.8%; Score 884; DB 2; Length 195;  
Best Local Similarity 83.6%; Pred. No. 4.1e-64;  
Matches 163; Conservative 18; Mismatches 14; Indels 0; Gaps 0;  
QY 1 VTPTLLSAFWTYLTITAMTPGPNNILALSSATSHGPROSTRVLAGMSLGFLLVMLLCAGI 60  
Db 1 MTPMLLSAFWYTYLTALTTPGPNNILALSAATAHGFRQSIKRVLAGMSLGFLLVMLLCAGI 60  
QY 61 SFSLAVIDPAVAHLLSWAGAAYIVWLAWKIATSPTEKEDGLQAKPISFWASFALQFVN VKI 120  
Db 61 AFSLAVIDPAIITHLSWGAAYIILWLAWKIATSPAENARPKPVGFVWSFGLQFVN VKI 120  
QY 121 ILXGVTALSTFVLPQTOALSWVGVSVLLAMIGTFGNVCWALAGHLFQRLFRQYGRQ LNI 180  
Db 121 ILXGITALSTFVLPQTOALNWVIGVSILLALIGTFGNVCWALAGHLFQRAFRHYGRQ LNI 180  
QY 181 VIALLLVYCAVRIFY 195  
Db 181 ILALLLVYCAVRIFY 195

RESULT 7  
Q7P2H4 PRELIMINARY; PRT; 200 AA.  
ID Q7P2H4

AC Q7P2H4;  
DT 01-MAR-2004 (Tremblrel. 26, Created)  
DT 01-MAR-2004 (Tremblrel. 26, Last sequence update)  
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)  
DE Transporter, Lyse family.  
GN Name=FNV0676;  
OS Fusobacterium nucleatum subsp. vincentii ATCC 49256.  
OC Bacteria; Fusobacteriales; Fusobacteriaceae;  
OC Fusobacterium.  
OX NCBI\_TaxID=209882;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 49256;  
RA Karpatral V., Ivanova N., Anderson I., Reznik G., Bhattacharyya A.,  
RA Gardner W.L., Mikhailova N., Larsen N., D'Souza M., Walunas T.,  
RA Haselkorn R., Overbeek R., Kyrpides N.;  
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.  
CC -! CAUTION: The sequence shown here is derived from an  
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
CC preliminary data.  
DR EMBL; AABF0100094; EAA23701.1; -.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0005293; F:lysine permease activity; IEA.  
DR GO; GO:0006865; P:amino acid transport; IEA.  
DR InterPro; IPR001123; Lyse.  
DR Pfam; PF01810; Lyse; 1.  
SQ SEQUENCE 200 AA; 22782 MW; 463825149BE6DA3 CRC64;

Query Match 36.9%; Score 367.5; DB 2; Length 200;  
Best Local Similarity 40.8%; Pred. No. 3.6e-22;  
Matches 78; Conservative 37; Mismatches 67; Indels 9; Gaps 3;

QY 5 LLSAFWYTYLTITAMTPGPNNILALSSATSHGPROSTRVLAGMSLGFLLVMLLCAGISFSL 64  
Db 8 MLFEYLPYALITNFTPGPNNILALNSTKYGFKRSWKVLLGVCLGFTICIMICIVCISL 67  
QY 65 AVIDPAVAHLLSWAGAAYIVWLAWKIATS-PTKEDGLQAKPISFWASFALQFVN VKIILY 123  
Db 68 KKVSDTYQNMKXIGALYIFWLAWHIFKSPSSILENSPKELTFLYGFIAQFVN VKIMLY 127  
QY 124 GVTALSTFVLPQTOA---LSWVGVSVLLAMIGTFGNVCWALAGHLFQRLFRQYGRQ LNI 179  
Db 128 GMVSISTFILPYQSALFIFLFI FGMSIL---GAIALVWALAGSLFQDFLNRHYRI FN 183  
QY 180 IVALLLVYCA 190  
Db 184 TVMGLILVKSA 194

RESULT 8  
Q63BD0 PRELIMINARY; PRT; 194 AA.  
ID Q63BD0  
AC Q63BD0;  
DT 25-OCT-2004 (Tremblrel. 28, Created)  
DT 25-OCT-2004 (Tremblrel. 28, Last sequence update)  
DT 25-OCT-2004 (Tremblrel. 28, Last annotation update)  
DE Transporter, Lyse family.  
GN Name=lyse; ORFNames=BTZK2196;  
OS Bacillus cereus ZK.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=288681;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ZK;  
RA Bretzin T.S., Bruce D., Challacombe J.F., Gilna P., Han C., Hill K.,  
RA Hitchcock P., Jackson P., Keim P., Longmire J., Lucas S., Okinaka R.,  
RA Richardson P., Rubin E., Tice H.;  
RT "Complete genome sequence of Bacillus cereus ZK."  
RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.  
DR EMBL; CP000001; AAU18061.1; -.  
SQ SEQUENCE 194 AA; 21905 MW; 2600626310681709 CRC64;

Query Match 31.8%; Score 316.5; DB 2; Length 194;



Best Local Similarity 36.4%; Pred. No. 4.9e-18;  
Matches 68; Conservative 40; Mismatches 78; Indels 1; Gaps 1;  
QY 8 AFWTYTLITAMTPGPNNILALSSATSHGFRQSTRVLAGMSLGLIYVLLCAGISFSLAVI 67  
Db 5 SFLLFVFISSTPGPNNFLAMTYANQHGLKSMQFCFVAFGFILTSLCSFNIILIKI 64  
QY 68 DPAVHLLSWAGAYIVWLAWKIATSPTKED-GLQAKPISFWASFALQFVNKIIYGV 126  
Db 65 LPIIEFPLKILGVAYMLYLAFKILTSKTSTDPDEKYNKNLFTVGILQFINPKGILFGLT 124  
QY 127 ALSTFVLPQTQALSWVGVSVLLAMIGTFGNVCWALAGHLFQRLFRQYGRQLNIYVALL 186  
Db 125 VVSTFILPYNSYSSYLLFSLFLGVGLMSTCSWSLFGSIFQKLLKHTKSFNIIMAVLL 184  
QY 187 VYCAVRI 193  
Db 185 VCSAISI 191  
RESULT 9  
Q81QH8 PRELIMINARY; PRT; 194 AA.  
AC Q81QH8; Q6HYP9; Q6KSQ2;  
DT 01-JUN-2003 (Tremblrel. 24, Created)  
DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)  
DT 25-OCT-2004 (Tremblrel. 28, Last annotation update)  
DE Transporter, lyse family.  
GN OrderedLocusNames=BA2447, BAS2278, GBAA2447;  
OS Bacillus anthracis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1392;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Ames / isolate Porton;  
RX MEDLINE=22608414; PubMed=12721629; DOI=10.1038/nature01586;  
RA Read T.D., Peterson S.N., Tourasse N.J., Baillye L.W., Paulsen I.T.,  
RA Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,  
RA Holtzaple E.K., Okstad O.A., Helgason B., Ristone J., Wu M.,  
RA Kolonay J.F., Beanan M.J., Dodson R.J., Brinkac L.M., Gwinn M.L.,  
RA DeBoy R.T., Madpu R., Daugherty S.C., Durkin A.S., Haft D.H.,  
RA Nelson W.C., Peterson J.D., Pop M., Khouiri H.M., Radune D.,  
RA Benton J.L., Mahamoud Y., Jiang L., Hance I.R., Weidman J.F.,  
RA Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Niernan W.C.,  
RA Hazen A., Cline R.T., Redmond C., Thwaite J.E., White O.,  
RA Salzberg S.L., Thomason B., Friedlander A.M., Koehler T.M.,  
RA Hanna P.C., Kolstoe A.-B., Fraser C.M.;  
RT "The genome sequence of Bacillus anthracis Ames and comparison to  
RT closely related bacteria.";  
RL Nature 423:81-86(2003).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Ames / isolate 0581;  
RA Ravel J., Rasko D.A., Shumway M.F., Jiang L., Cer R.Z., Federova N.B.,  
RA Wilson M., Stanley S., Decker S., Read T.D., Salzberg S.L.,  
RA Fraser C.M.;  
RT "Bacillus anthracis comparative genomics.";  
RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Sterne;  
RA Brettin T.S., Bruce D., Challacombe J.F., Gilna P., Han C., Hill K.,  
RA Hitchcock P., Jackson P., Keim P., Longmire J., Lucas S., Okinaka R.,  
RA Richardson P., Rubin E., Tice H.;  
RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AE017031; AAP26309.1; -  
DR EMBL; AE017334; AAT31562.1; -  
DR EMBL; AE017225; AAT54590.1; -  
DR TIGR; BA2447; -  
DR TIGR; GBAA2447; -  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0005293; F:lysine permease activity; IEA.  
DR GO; GO:0006865; P:amino acid transport; IEA.

DR InterPro; IPR001123; Lyse.  
DR Pfam; PF01810; Lyse; 1.  
KW Complete proteome.  
SQ SEQUENCE 194 AA; 21820 MW; E1065D342BDE575E CRC64;  
Query Match  
Best Local Similarity 31.3%; Score 311.5; DB 2; Length 194;  
Matches 67; Conservative 40; Mismatches 79; Indels 1; Gaps 1;  
QY 8 AFWTYTLITAMTPGPNNILALSSATSHGFRQSTRVLAGMSLGLIYVLLCAGISFSLAVI 67  
Db 5 SFLLFVFISSTPGPNNFLAMTYANQHGLKSMQFCFVAFGFILTSLCSFNIILIKI 64  
QY 68 DPAVHLLSWAGAYIVWLAWKIATSPTKED-GLQAKPISFWASFALQFVNKIIYGV 126  
Db 65 LPIIEFPLKILGVAYMLYLAFKILTSKTSTDPDEKYNKNLFTVGILQFINPKGILFGLT 124  
QY 127 ALSTFVLPQTQALSWVGVSVLLAMIGTFGNVCWALAGHLFQRLFRQYGRQLNIYVALL 186  
Db 125 VVSTFILPYNSYSSYLLFSLFLGVGLMSTCSWSLFGSIFQKLLKHTKSFNIIMAVLL 184  
QY 187 VYCAVRI 193  
Db 185 VCSAISI 191  
RESULT 10  
Q6HIR1 PRELIMINARY; PRT; 194 AA.  
AC Q6HIR1;  
DT 05-JUL-2004 (Tremblrel. 27, Created)  
DT 05-JUL-2004 (Tremblrel. 27, Last sequence update)  
DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)  
DE Transporter, lyse family.  
GN Name=lyse; OrderedLocusNames=BT9727\_2238;  
OS Bacillus thuringiensis (subsp. konkukian).  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=180856;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=97-27;  
RA Brettin T.S., Bruce D., Challacombe J.F., Gilna P., Han C., Hill K.,  
RA Hitchcock P., Jackson P., Keim P., Longmire J., Lucas S., Okinaka R.,  
RA Richardson P., Rubin E., Tice H.;  
RT "Complete genome sequence of Bacillus thuringiensis 97-27.";  
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AE017355; AAT59905.1; -  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0005293; F:lysine permease activity; IEA.  
DR GO; GO:0006865; P:amino acid transport; IEA.  
DR InterPro; IPR001123; Lyse.  
DR Pfam; PF01810; Lyse; 1.  
KW Complete proteome.  
SQ SEQUENCE 194 AA; 21820 MW; E1065D342BDE575E CRC64;  
Query Match  
Best Local Similarity 31.3%; Score 311.5; DB 2; Length 194;  
Matches 67; Conservative 40; Mismatches 79; Indels 1; Gaps 1;  
QY 8 AFWTYTLITAMTPGPNNILALSSATSHGFRQSTRVLAGMSLGLIYVLLCAGISFSLAVI 67  
Db 5 SFLLFVFISSTPGPNNFLAMTYANQHGLKSMQFCFVAFGFILTSLCSFNIILIKI 64  
QY 68 DPAVHLLSWAGAYIVWLAWKIATSPTKED-GLQAKPISFWASFALQFVNKIIYGV 126  
Db 65 LPIIEFPLKILGVAYMLYLAFKILTSKTSTDPDEKYNKNLFTVGILQFINPKGILFGLT 124  
QY 127 ALSTFVLPQTQALSWVGVSVLLAMIGTFGNVCWALAGHLFQRLFRQYGRQLNIYVALL 186  
Db 125 VVSTFILPYNSYSSYLLFSLFLGVGLMSTCSWSLFGSIFQKLLKHTKSFNIIMAVLL 184  
QY 187 VYCAVRI 193  
Db 185 VCSAISI 191



Db 185 VCSAISI 191

RESULT 11

Q738B5 PRELIMINARY; PRT; 194 AA.

AC Q738B5;

DT 05-JUL-2004 (TReMBLrel. 27, Created)

DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)

DT 05-JUL-2004 (TReMBLrel. 27, Last annotation update)

DE Transporter, Lyse family.

GN OrderedLocustNames=BCE2480;

OS Bacillus cereus (strain ATCC 10987).

OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

OX NCBI\_TaxID=222523;

RM [1]

RP SEQUENCE FROM N.A.

RX PubMed=14960714; DOI=10.1093/nar/gkh258;

RA Rasako D.A., Ravel J., Oekstad O.A., Helgason E., Cer R.Z., Jiang L., Shores K.A., Fouts D.E., Tourasse N.J., Angluoli S.V., Kolonay J.F., Nelson W.C., Kolstoe A.-B., Fraser C.M., Read T.D.;

RA "The genome sequence of Bacillus cereus ATCC 10987 reveals metabolic adaptations and a large plasmid related to Bacillus anthracis pXOI.";

RL Nucleic Acids Res. 32:977-988(2004).

DR EMBL; AE017272; AAS41397.1; -.

DR TIGR; BCE2480; -.

DR GO; GO:0016020; C:membrane; IEA.

DR GO; GO:0005293; F:Lysine permease activity; IEA.

DR GO; GO:0006865; P:amino acid transport; IEA.

DR InterPro; IPR001123; Lyse.

DR Pfam; PF01810; Lyse; 1.

KW Complete proteome.

SQ SEQUENCE 194 AA; 21828 MW; D4B92C4E54BE2E26 CRC64;

Query Match 31.2%; Score 310.5; DB 2; Length 194;

Best Local Similarity 36.4%; Pred. No. 1.5e-17;

Matches 68; Conservative 39; Mismatches 79; Indels 1; Gaps 1;

QY 8 AFWTYTLITAMTPGPNNIILASSATSHGFRQSTRVLGMSLGLIVMLCAGISFSLAVI 67

Db 5 SFLLFVITCSFTPGPNFLAMTYANQHGLKSMQFCFVAFGFILLTSLCSFFNIVLINI 64

QY 68 DPAVHLTWSGAAIYIWLAWKIATSPTEKEDGLQAKPIS-FWASFAIQFVNKIIYGV 126

Db 65 LPIIEFPLKILGVAYMLYLAFKILTSKTSADPDEKHNKULFTVGIFLQFVNPKIILFGLT 124

QY 127 ALSTFVLPTQALSWVGVSVLLAMIGTFGNVCWALAGHLFQRLFRQYGRQNLIVLALL 186

Db 125 VVSTFILPYNYSFLLFSLFLGVGLMSTFWSLFGSMFQKFLKHTTSFNITMAVLL 184

QY 187 VYCAVRI 193

Db 185 VFS AISI 191

RESULT 12

Q81D17 PRELIMINARY; PRT; 194 AA.

AC Q81D17;

DT 01-JUN-2003 (TReMBLrel. 24, Created)

DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)

DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)

DE Transporter, Lyse family.

GN OrderedLocustNames=BC2378;

OS Bacillus cereus (strain ATCC 14579 / DSM 31).

OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

OX NCBI\_TaxID=226900;

RM [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=22608415; PubMed=12721630; DOI=10.1038/nature01582;

RA Ivanova N., Sorokin A., Anderson I., Galleron N., Candelson B., Kapatai V., Bhattacharyya A., Reznik G., Mikhailova N., Lapidus A., Chu L., Mazur M., Goltsman E., Larsen N., D'Souza M., Walunas T.,

RA Grechkin Y., Pusch G., Haselkorn R., Fonstein M., Ehrlich S.D.,

RA Overbeek R., Kyrpides N.C.;

RT "Genome sequence of Bacillus cereus and comparative analysis with

RT Bacillus anthracis.";

RL Nature 423:87-91(2003).

DR EMBL; AE017005; AAP09341.1; -.

DR GO; GO:0016020; C:membrane; IEA.

DR GO; GO:0005293; F:Lysine permease activity; IEA.

DR GO; GO:0006865; P:amino acid transport; IEA.

DR InterPro; IPR001123; Lyse.

DR Pfam; PF01810; Lyse; 1.

KW Complete proteome.

SQ SEQUENCE 194 AA; 21777 MW; 393B66AF87C8C96F CRC64;

Query Match 30.5%; Score 303.5; DB 2; Length 194;

Best Local Similarity 36.2%; Pred. No. 5.5e-17;

Matches 68; Conservative 39; Mismatches 78; Indels 3; Gaps 2;

QY 8 AFWTYTLITAMTPGPNNIILASSATSHGFRQSTRVLGMSLGLIVMLCAGISFSLAVI 67

Db 5 SFLLFVITCSFTPGPNFLAMTYANQHGLKRSINFCFVAFGFILLTSLCSFFNIVLINI 64

QY 68 DPAVHLTWSGAAIYIWLAWKIATSP--TKEDGLQAKPISFWASFAIQFVNKIIYGV 125

Db 65 LPIIEFPLKILGVAYMLYLAFKIITSKGAGHNEKNKNL-FTVGIFLQFINPKIILFGL 123

QY 126 TALSTFVLPTQALSWVGVSVLLAMIGTFGNVCWALAGHLFQRLFRQYGRQNLIVLALL 185

Db 124 TVVSTFILPYNYSFLLFSLFLGIVGLSTFWSLFGSMFQKLLKHNQLEFNITMAVL 183

QY 186 LVYCAVRI 193

Db 184 LVFS AVSI 191

RESULT 13

Q65N28 PRELIMINARY; PRT; 192 AA.

AC Q65N28;

DT 25-OCT-2004 (TReMBLrel. 28, Created)

DT 25-OCT-2004 (TReMBLrel. 28, Last sequence update)

DT 25-OCT-2004 (TReMBLrel. 28, Last annotation update)

DE Lyse (lysine exporter protein) (LYSE/YGGA).

GN Name=lyse; ORFNames=BL02014, BLi00587;

OS Bacillus licheniformis DSM 13.

OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

OX NCBI\_TaxID=279010;

RM [1]

RP SEQUENCE FROM N.A.

RC STRAIN=DSM 13;

RX PubMed=15383718;

RA Veith B., Herzberg C., Steckel S., Feesche J., Maurer K.H., Ehrenreich P., Baumer S., Henne A., Liesegang H., Merkl R.,

RA "The Complete Genome Sequence of Bacillus licheniformis DSM13, an Organism with Great Industrial Potential.";

RT J. Mol. Microbiol. Biotechnol. 7:204-211(2004).

RL [2]

RM SEQUENCE FROM N.A.

RC STRAIN=ATCC 14580;

RA Tang M., de Leon A.L., Xiang H., Gusti V., Clausen I.G., Olsen P.B., Rasmussen M.D., Andersen J.T., Jorgensen P.J., Larsen T.S., Sorokin A., Bolotin A., Lapidus A., Galleron N., Ehrlich S.D., Berka R.M.;

RT "Complete genome sequence of the industrial bacterium Bacillus licheniformis and comparisons with closely related Bacillus species.";

RL Genome Biol. 5:R77-R77(2004).

DR EMBL; AE017333; AAU39536.1; -.

DR EMBL; CP000002; AAU22181.1; -.

SQ SEQUENCE 192 AA; 21542 MW; 43BDD25276C074EE CRC64;

Query Match 29.4%; Score 293; DB 2; Length 192;



*This Page Blank (uspto)*